

# Further empirical evidence for the non-linearity of the period-luminosity relations as seen in the Large Magellanic Cloud Cepheids

Chow-Choong Ngeow<sup>1,5\*</sup>, Shashi M. Kanbur<sup>1,6</sup>, Sergei Nikolaev<sup>2</sup>, John Buonaccorsi<sup>3</sup>,  
Kem H. Cook<sup>2</sup> and Douglas L. Welch<sup>4</sup>

<sup>1</sup>*Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA*

<sup>2</sup>*Institute for Geophysics and Planetary Physics, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA*

<sup>3</sup>*Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003, USA*

<sup>4</sup>*Department of Physics and Astronomy, McMaster University, Hamilton, ON L8S 4M1, Canada*

<sup>5</sup>*Department of Astronomy, University of Illinois, Urbana-Champaign, IL 61801, USA*

<sup>6</sup>*Department of Physics, State University of New York at Oswego, Oswego, NY 13126, USA*

Accepted 2005 month day. Received 2005 month day; in original form 2005 April. 11

## ABSTRACT

Recent studies, using OGLE data for LMC Cepheids in the optical, strongly suggest that the period-luminosity (PL) relation for the Large Magellanic Cloud (LMC) Cepheids shows a break or non-linearity at a period of 10 days. In this paper we apply statistical tests, the chi-square test and the  $F$ -test, to the Cepheid data from the MACHO project to test for a non-linearity of the  $V$ - and  $R$ -band PL relations at 10 days, and extend these tests to the near infrared ( $JHK$ -band) PL relations with 2MASS data. We correct the extinction for these data by applying an extinction map towards the LMC. The statistical test we use, the  $F$ -test, is able to take account of small numbers of data points and the nature of that data on either side of the period cut at 10 days. With our data, the results we obtained imply that the  $VRJH$ -band PL relations are non-linear around a period of 10 days, while the  $K$ -band PL relation is (marginally) consistent with a single-line regression. The choice of a period of 10 days, around which this non-linearity occurs, is consistent with the results obtained when this "break" period is estimated from the data.

We show that robust parametric (including least squares, least absolute deviation, robust regression) and non-parametric regression methods, which restrict the influence of outliers, produce similar results. Long period Cepheids are supplemented from the literature to increase our sample size. The photometry of these long period Cepheids is compared with our data and no trend with period is found. Our main results remain unchanged when we supplement our dataset with these long period Cepheids. By examining our data at maximum light, we also suggest arguments why errors in reddening are unlikely to be responsible for our results. The non-linearity of the mean  $V$ -band PL relation as seen in both of the OGLE and MACHO data, using different extinction maps, suggests that this non-linearity is real.

**Key words:** Cepheids – Stars: fundamental parameters

## 1 INTRODUCTION

The Cepheid period-luminosity (PL) relation plays a crucial role in the distance ladder which can ultimately be

used to determine Hubble constant. The current most widely used PL relation is based on the Large Magellanic Cloud (LMC) Cepheids (e.g., see Freedman et al. 2001; Saha et al. 2001). The Cepheid PL relation has long been considered to be a linear function of  $\log(P)$  within the range of  $\log(P) \sim 0.3$  to  $\log(P) \sim 2.0$ , where  $P$  is the

\* E-mail: cngeow@astro.uiuc.edu

pulsation period in days. Current versions of the LMC PL relations can be found, e.g., in Madore & Freedman (1991), Tanvir (1997), Gieren et al. (1998), Udalski et al. (1999a) and Persson et al. (2004).

However, recent studies (Tammann & Reindl 2002; Tammann et al. 2002; Kanbur & Ngeow 2004; Sandage et al. 2004) strongly suggest that the LMC PL relation is not linear, as there are two PL relations for the short ( $P < 10$  days) and the long ( $P > 10$  days) period LMC Cepheids, respectively. These studies are mainly based on the optical data (*BVI*-band) from the Optical Gravitational Lensing Experiment (OGLE, Udalski et al. 1999a,b) database for the fundamental mode Cepheids. The non-linear *V*- and *I*-band LMC PL relations are further supported by the results from a rigorous statistical test, the *F*-test, as described in Kanbur & Ngeow (2004, hereafter KN). This *F*-test is sensitive to the number of data points and the nature of the data on either side of 10 days. This sensitivity is such that if, either the number of data points is small or the data has a large scatter, the slopes on either side of 10 days are not well defined and the test returns a non-significant result. We describe this property of the *F*-test in detail in Section 3. Hence the work of KN, which used the OGLE data alone and found marked non-linearities in the PL (and PC, period-colour) relations in the LMC at a period of 10 days is in contrast to Gieren et al. (2005), who state that the OGLE data *alone* do not show any non-linear behavior or break at a period of 10 days. Since the OGLE Cepheid data are truncated at  $\log(P) \sim 1.5$  (Udalski et al. 1999a), Sandage et al. (2004) used additional *BVI*-band data that are available from the literature, especially those with  $\log(P) > 1.5$ , to further support the existence of two PL relations in the LMC Cepheids. Compared to previous studies, the large number of LMC Cepheids with high quality photometric data, together with the estimation of extinction ( $E[B - V]$ ) to individual Cepheids, given in the OGLE database has made the detection of non-linear LMC PL relation becomes possible.

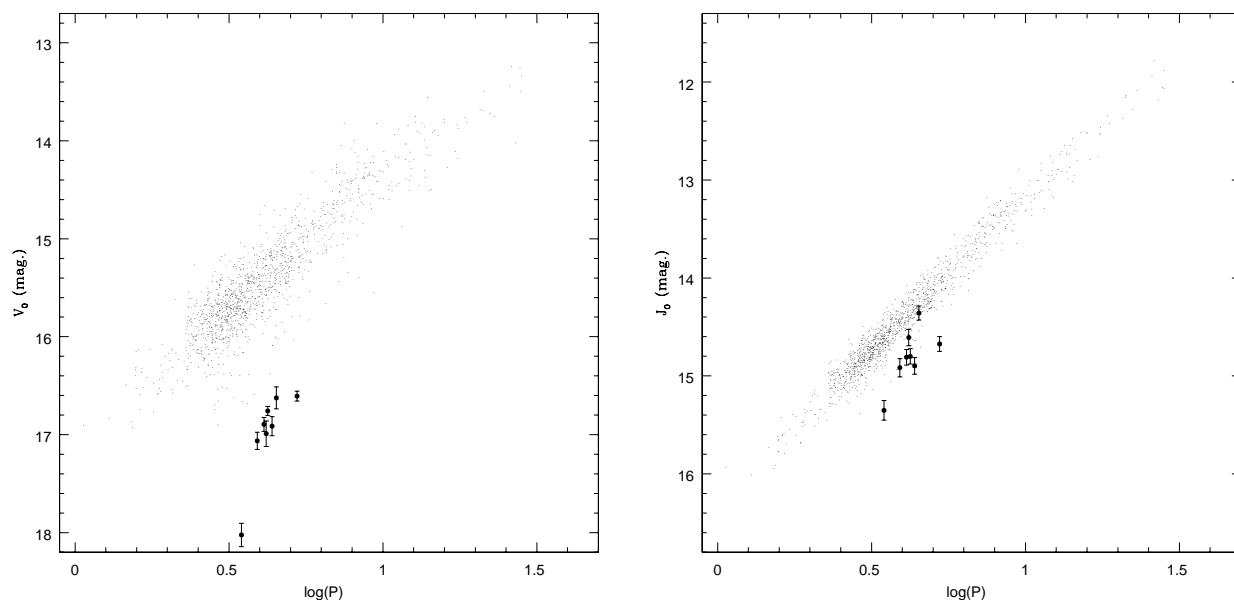
The non-linearity of Cepheid PL relations has been recognized for some time in the literature, as the mean magnitudes for some of the long period Cepheids ( $\log[P] \gtrsim 1.5$ ) do not follow the best-fit lines. In fact, some of the earlier studies (including the works of Kukarkin, Shapley & Fernie) on the Cepheid PL relation have suggested or used a quadratic form,  $M = \alpha + \beta \log(P) + \gamma [\log(P)]^2$ , to describe the PL relation (Fernie 1969). For example, a semi-empirical derivation of the quadratic PL relations in the *B*- and *V*-band can be found in Fernie (1967). In addition, the composite PL relations at mean and maximum light constructed with Cepheids in different galaxies, as presented in Sandage & Tammann (1968), also show some curvature of the PL relations. In terms of theoretical modeling, Bono et al. (1999), Caputo et al. (2000) and Marconi et al. (2005) have fit quadratic PL relations to the theoretical periods and magnitudes obtained from the pulsational modeling. The non-linearity of the PL relations, however, can also be explained with the broken PL relations: there are two PL relations, for short and long period Cepheids respectively, with a break at a specific

period (e.g., at 10 days). Perhaps the earliest idea for the broken PL relation was proposed by Kukarkin in 1937, as quoted from Fernie (1969): “...he concluded that the relation was definitely non-linear...but the best fit was by two straight lines of different slope intersecting at a period of 10 days”. The present-day version of the broken PL relations is presented in Tammann & Reindl (2002), KN and Sandage et al. (2004) with the OGLE data, as mentioned in the previous paragraph. These broken PL relation are also compared with the theoretical predictions in Marconi et al. (2005), which show a quite satisfactory agreement. Note that Caputo et al. (1999) have fit a broken PL relation to their results obtained from the theoretical models, but with a break at  $\log(P) = 1.4$ .

Besides the non-linear PL relations seen in the LMC Cepheids, the SMC fundamental mode Cepheids also show a change of slope for Cepheids with periods less than 2 days from the EROS-2 data (Bauer et al. 1999). The authors carefully examine the possible errors or bias in the data that lead to this finding, and conclude that the non-linear PL relations seen in SMC Cepheids is real. Baraffe et al. (1998) explained that this change of the slope for short period (less than 2 days) SMC Cepheids is due to evolutionary effects: this is because the blue loops for these short period SMC Cepheids do not fully extend into the instability strip. However, the non-linearity of the LMC PL relation at  $\sim 10$  days period is believed to be independent of this evolutionary effect. In this paper, we concentrate on studying the 10 days non-linearity for the LMC PL relation and do not suggest that this is the same phenomena described by Baraffe et al. (1998) but for the higher metallicity LMC Cepheids.

The motivation of this paper is to extend the work of KN into the study of the linear and/or non-linear nature of the PL relations at/around 10 days for the fundamental mode LMC Cepheids. It is important to verify the non-linearity of the LMC PL relation at 10 days from the studies of Tammann & Reindl (2002), KN and Sandage et al. (2004), which mainly analyzed the OGLE data, with an independent dataset. Hence, we examine the non-linearity of the optical LMC PL relation with the Cepheid data obtained from the MACHO database (Alcock et al. 2000) in this paper. In addition, we also investigate the LMC PL relations in the near infrared (NIR) bands (i.e., *JHK*-band) to see if the NIR PL relations also show a non-linearity. We apply statistical tests to investigate the possible non-linearity of the LMC PL relation in these bands. We formulate this non-linearity in three ways and examine if the data are more consistent with a linear or non-linear PL relation.

In this paper, we only use the published mean magnitudes (and periods) for the LMC Cepheids from the literature. The data used in this paper is described in Section 2. The statistical analysis and the results are presented in Section 3 and 4, respectively. The conclusion & discussion will be followed in Section 5.



**Figure 1.** The extinction corrected  $V$ - and  $J$ -band PL relations with all 1330 Cepheids in the sample. The filled circles are the 8 outliers as described in the text. The error bars for the outliers are the quadrature sum of the photometric errors and the extinction errors. The error bars for other Cepheids are omitted for clarity.

## 2 DATA AND EXTINCTION CORRECTIONS

The data for the 1330 fundamental mode LMC Cepheids are adopted from Nikolaev et al. (2004). This includes the  $V$ - and  $R$ -band mean magnitudes, as well as the periods, from the MACHO database and the  $J$ -,  $H$ - and  $K$ -band mean magnitudes from the Two Micron All Sky Survey database (2MASS, Skrutskie 1998). The mean magnitudes in the  $V$ - and  $R$ -bands are computed from the complete light curves with the MACHO photometric data. The (random phase) magnitudes in  $J$ -,  $H$ - and  $K$ -bands are from the 2MASS single epoch observations, hence the mean magnitudes in these NIR bands are obtained with a random-phase correction function. The details for obtaining the mean magnitudes in all these bands can be found in Nikolaev et al. (2004). To check the accuracy of these  $JHK$  mean magnitudes, one of us (Nikolaev et al, 2005 – in preparation) has compared the  $JHK$  mean magnitudes from the 2MASS data (after random-phase correction) with the Persson et al. (2004) data, which are obtained from the well-sampled ( $JHK$ ) light curves. For the 33 matched Cepheids in the MACHO/2MASS sample and the Persson et al. (2004) sample, the average difference in the mean magnitudes are:  $\overline{\Delta_J} = -0.022$  (with  $\text{std} = 0.070$ , where  $\text{std}$  = standard deviation),  $\overline{\Delta_H} = -0.012$  ( $\text{std} = 0.073$ ) and  $\overline{\Delta_K} = -0.015$  ( $\text{std} = 0.068$ ). These are completely consistent with the zero difference. More importantly, the plots of the difference between the means in these bands as a function of period (not shown) do not show any trends with period, hence the  $JHK$  mean magnitudes from the 2MASS database will not seriously affect our analysis and results. We did not use other datasets in  $VRJHK$ -bands, as in Sandage et al. (2004), to perform the main analysis in order to keep our dataset as consis-

tent and homogeneous as possible (however, see Section 5.2).

The next step in the analysis is to obtain the extinction-corrected mean magnitudes for these Cepheids. This was done by applying an extinction map with the following values of  $R_\lambda$  (ratio of total-to-selective reddening, and  $R_\lambda = A_\lambda/E[B-V]$ ), in order to be consistent with Nikolaev et al. (2004):  $R_{V,R,J,H,K} = \{3.12, 2.56, 0.90, 0.53, 0.34\}$ . The LMC extinction map we used is taken from Zaritsky et al. (2004). In brief, given the coordinates for each Cepheid and a search radius  $r$ , the extinction map returns the average value of extinction,  $\overline{A_V} \pm \sigma_{A_V}$  (where  $\sigma_{A_V} = \text{std}/\sqrt{N}$ ), from the stars bounded within the search radius. The extinction map allows us to choose the  $\overline{A_V}$  obtained from cool stars ( $5500\text{K} < T_{\text{eff}} < 6500\text{K}$ ), hot stars ( $12,000\text{K} < T_{\text{eff}} < 45,000\text{K}$ ) or both. Since Cepheid variables are cool stars, we only use the extinction values from the cool stars. We use  $r = 2'$  to ensure that there are sufficient numbers of stars within the search radius to obtain the  $\overline{A_V}$  (usually more than 10 stars)<sup>1</sup>. However, there are two Cepheids with locations beyond the extinction map and one Cepheid without any cool stars within the  $r = 3'$  radius. For simplicity, we assume  $\overline{A_V} = 0.3 \pm 0.1$  mag (roughly the mean extinction of LMC) for these three Cepheids. After the values of  $\overline{A_V}$  for the Cepheids are obtained, the extinctions in other bands are calculated with the following relation:  $A_\lambda = (\frac{R_\lambda}{R_V})\overline{A_V}$  with the errors  $\sigma_{A_\lambda} = (\frac{R_\lambda}{R_V})\sigma_{A_V}$ , where  $\lambda$  denotes the passband.

We consider the following three samples in this study:

<sup>1</sup> Only two Cepheids with enclosed stars that are less than 10

**Sample A** - This sample consist of all 1330 Cepheids in our dataset.

**Sample B** - We then remove 8 outliers out of 1330 Cepheids (all of these outliers have  $\log[P] < 0.75$ ) from Sample A, as they are fainter by more than  $\sim 1.2\text{mag}$ . and  $\sim 1.0\text{mag}$ . in  $V$ - and  $R$ -band, respectively, to the fitted PL relations at the same period. The locations of these outliers in the PL relations are plotted in Figure 1. It is possible that the extinctions for these outliers are underestimated because they are not obvious in the NIR PL relations (see the right panel of Figure 1). Another possibility is that they are Type II Cepheids that are mis-identified as the classical Type I Cepheids. However, it is clear that these outliers should be removed from our sample. Hence Sample B consists of 1322 Cepheids.

**Sample C** - Following Udalski et al. (1999a), we also remove the Cepheids with  $\log(P) < 0.4$  from Sample B. This is to guard against the possible contamination from first overtone Cepheids (Udalski et al. 1999a), as well as other types of variables such as the anomalous Cepheids and the double-mode Cepheids (Alcock et al. 1999). In addition, this choice of the period cut will also guard against the possible change of slopes for short ( $P < 2$  days) period Cepheids, as seen in the SMC Cepheids (Bauer et al. 1999; Udalski et al. 1999a). Note that Udalski et al. (1999a) has applied the period cut of  $\log(P) = 0.4$  to the OGLE LMC Cepheids to derive the PL relations that were used in other studies, such as the  $H_0$  Key project (Freedman et al. 2001). Hence the same period cut was also applied in previous studies of the non-linear LMC PL relations (Tammann & Reindl 2002; Kanbur & Ngeow 2004; Sandage et al. 2004). Furthermore, the period distribution of the LMC Cepheids shows a sharp break at  $\log(P) \sim 0.4$  (Alcock et al. 1999). This is probably due to evolutionary effects (Baraffe et al. 1998; Alcock et al. 1999). We contend that these reasons justify the removal of Cepheids with  $\log(P) < 0.4$ , and leaves 1216 Cepheids in this sample. Our use of the  $F$ -test, as given in the next section, ensures that such a period cut will not affect our results.

### 3 ANALYSIS

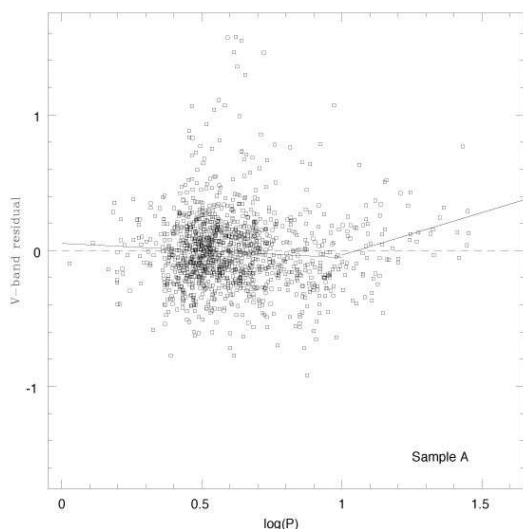
In order to study the linear and/or non-linear nature of the LMC PL relation, the data presented in the previous section are fitted with the (weighted) least squares regression (see, e.g., Press et al. 1992). Ordinary least squares regression,  $OLS(Y|X)$ , has been extensively used in the literature to obtain the fitted PL relations from observations (for examples, see Madore & Freedman 1991; Laney & Stobie 1994; Gieren et al. 1998; Udalski et al. 1999a; Kanbur & Ngeow 2004; Sandage et al. 2004; Freedman et al. 2001, and many other papers on PL relation using the Galactic Cepheids) as well as from theoretical modelings (as in Bono et al. 1999; Caputo et al. 2000; Baraffe & Alibert 2001). This is because there is a good physical theory (see, e.g., Sandage 1958; Cox 1980; Madore & Freedman 1991) suggesting why the period is the independent variable and is measured essentially without errors, hence the  $OLS(Y|X)$  regression can be applied. In terms of observations, the errors of the

periods for the MACHO Cepheids used in this paper (see Section 2) is about  $10^{-4}P$  (Nikolaev et al. 2004), which is comparable to the accuracy reported by the OGLE team ( $\sigma_P \sim 10^{-5}P$ , Udalski et al. 1999b), as well as other high quality photometric observations (e.g., see Moffett et al. 1998, and some examples from the Galactic Cepheids in Coulson et al. 1985, Shobbrook 1992 and Bersier et al. 1994). The  $OLS(Y|X)$  regression is also recommended for the distance scale studies (Isobe et al. 1990; Feigelson & Babu 1992).

The weights ( $w$ ) that are used in the regressions are the quadrature sum of the photometric errors in mean magnitudes ( $\sigma_\lambda$ ), the errors from the random phase corrections that are only applicable to  $JHK$ -bands ( $\sigma_{rp}$ ), the errors from extinction ( $\sigma_{A_\lambda}$ ), and the intrinsic dispersion due to the finite width of the instability strip ( $\sigma_{IS}$ ). Hence,  $w(i) = \sigma_\lambda^2(i) + \sigma_{rp}^2(i) + \sigma_{A_\lambda}^2(i) + \sigma_{IS}^2$ . The dispersions of the PL relations in each band from Madore & Freedman (1991) can be initially adopted for  $\sigma_{IS}$ . However, a comparison of the dispersions given in Madore & Freedman (1991) with Gieren et al. (1998), Persson et al. (2004) and Sandage et al. (2004) suggest that the  $VR$ -band and the  $JHK$ -band dispersions given in Madore & Freedman (1991) should be scaled by  $\sim 0.73$  and  $\sim 0.88$ , respectively. Hence we adopt the following values for the intrinsic dispersions:  $\sigma_{IS}(V, R, J, H, K) = \{0.21, 0.18, 0.14, 0.12, 0.11\}$ . The adopted  $\sigma_{IS}$ , such as those originally given in Madore & Freedman (1991), does not affect the results from the following statistical tests. We also assume that all LMC Cepheids are at the same distance so that there is no distance variation in the weight formula.

One way to check for a non-linearity on the LMC PL relation is through a residual plot. If the one-line regression is adequate, then the residuals should be populated (roughly) evenly around the zeroth residual line (i.e., independent of period). On the other hand, if the residuals show a trend, or some trends, in the residual plot, then more parameters are needed to fit the data. The residuals of the MACHO  $V$ -band data from previous section are plotted as function of  $\log(P)$  in Figure 2, assuming that the one-line regression is adequate. A linear fit of the residuals with  $\log(P) > 1.0$  gives non-zero slopes of  $-0.66 \pm 0.27$ ,  $-0.65 \pm 0.27$  and  $-0.67 \pm 0.27$  from Sample A, B and C, respectively. These slopes clearly indicate that the trends seen in Figure 2 are significant. Hence a one-line regression may not be adequate to fit the MACHO  $V$ -band PL relations. Similar residual plots in other bands are presented in the right panels of Figure 4.

To investigate the possible non-linearity of the PL relations, we applied the chi-square test and the  $F$ -test (as given in Weisberg 1980, and used initially in KN) using the weighted least squares results. The null ( $H_O$ ) and the alternate ( $H_A$ ) hypothesis represent the reduced and the full models, respectively, where the reduced models are obtained by setting some parameters in the full models to certain specific values (e.g., zero, see Weisberg 1980). Hence our reduced model is a one-line regression (2-parameter) fit to the entire period range, and the null hypothesis,  $H_O$ , is that one-line regression is adequate to describe the data. In contrast,  $H_A$  is that the full model is



**Figure 2.** The residuals for the V-band PL relation as a function of log-period from Sample A. The residuals are obtained assuming the one-line regression is adequate. The solid lines are the linear fits of the residuals as function of log-period, separated at 10 days. The dashed line is the zeroth residual line. The residual plots from Sample B & C are similar to this plot.

necessary to fit the data. Here, we consider three different kinds of full models:

(i) A two-line regression (4-parameter) with a break at a characteristic period,  $P_o$ , i.e.,

$$\begin{aligned} m &= a^L \log(P) + b^L, P > P_o, \\ m &= a^S \log(P) + b^S, P < P_o. \end{aligned} \quad (1)$$

(ii) A piecewise continuous regression (3-parameter) that forces the two-line regression to be continuous at  $P_o$ , i.e.,

$$\begin{aligned} m &= c + d \times \min[\log(P), \log(P_o)] \\ &\quad + e \times \max[0, \log(P) - \log(P_o)]. \end{aligned} \quad (2)$$

(iii) A quadratic regression (3-parameter) of the form of:

$$m = \alpha + \beta \log(P) + \gamma [\log(P)]^2. \quad (3)$$

Although we consider these three models, our main interest is in the investigation of a change of slope in the LMC PL relation for short and long period Cepheids. It is clear that two lines of differing slope, continuous at  $P_o$ , and even perhaps with a small discontinuity at  $P_o$ , are approximately a quadratic with a local minimum/maximum derivative at  $P_o$ .

To apply the chi-square test and the  $F$ -test, we then calculated the weighted error sum of squares,  $SSE = \sum [m(i) - \hat{m}(i)]^2 / w(i)$ , for both  $H_O$  and  $H_A$ , where  $m(i)$  and  $\hat{m}(i)$  are the observed magnitudes and the fitted magnitudes from weighted least squares for  $i$ th Cepheid, respectively. Under  $H_O$ , the difference of the  $SSE$ :

**Table 1.** The critical values for the chi-square test and the  $F$ -test.

$\nu^a$	$p = 0.05$	$p = 0.01$
$\chi^2$ values		
1	3.84	6.63
2	5.99	9.21
$F$ values		
1	3.84	6.63
2	3.00	4.61

<sup>a</sup>  $\nu = 1$  for the alternate hypothesis given in equation (2) & (3);  $\nu = 2$  for the alternate hypothesis given in equation (1).

$$\chi^2 = SSE(H_O) - SSE(H_A), \quad (4)$$

follows approximately a chi-square distribution with  $\nu = n_{FL} - n_{RD}$  degrees of freedom (e.g, see Weisberg 1980), where  $n_{FL}$  &  $n_{RD}$  are the number of parameters in full and reduced models, respectively. For a large number of data points ( $N$ ), this is essentially the same as applying a standard  $F$ -test (which many regression packages produce) under weighted least squares, where

$$F = \frac{\chi^2 / \nu}{SSE(H_A) / (N - n_{FL})}, \quad (5)$$

which is approximately a  $F_{(\nu, N - n_{FL})}$  distribution under the null hypothesis (see also Kanbur & Ngeow 2004), where  $\chi^2$  is given in equation (4). The corresponding probability ( $p$ ) for the  $\chi^2$  and the  $F$ -values, under  $H_O$ , can be obtained from consulting the standard statistical books or softwares. A large value of  $\chi^2$  and/or  $F$  (hence small  $p$ ) indicates that the null hypothesis can be rejected, and therefore the data is more consistent with the alternate hypothesis. Table 1 summarizes the  $\chi^2$  and  $F$ -values that correspond to the case of  $p = 0.05$  and  $p = 0.01$ . In case of the  $F$ -test,  $p(F) < 10^{-3}$  for  $F > 7$  with large number of data points (say,  $N > 500$ ).

The  $F$ -test relies on the assumptions that the residuals are normally distributed and are homoskedastic. KN tested the OGLE data for these assumptions. With such large numbers of data departures from normality are unlikely to significantly influence our results. The chi-square test, as formulated above, does not depend on homoskedasticity and hence we use it in this study to supplement our  $F$ -test results. Perhaps a simpler chi-square test is the following: compare the chi-square statistic for a one- and two-line regression respectively. However, this approach does not have the necessary power to distinguish between the two cases and hence we use the chi-square test given in equation (4) above.

Furthermore, the  $F$ -test we use, as in equation (5), to test for changes in slope on either side of a period cut, is sensitive to the number and nature of Cepheid data (such as the distribution of the data) defining the slope on either side of the period cut. Equation (5) contains an explicit dependence on these factors. Since the variance of the fitted slope,  $\hat{a}$ , is  $\text{Var}(\hat{a}) = \frac{1}{(N-n)} \frac{SSE}{S_{XX}}$ , where  $n$  is number of parameters and  $S_{XX} = \sum (x - \bar{x})^2$  ( $\bar{x}$  is the mean value of  $x = \log[P]$ ), hence another

**Table 2.** Testing the significance of the  $F$ -test with various period cuts.

$\log(P_1)$	$\log(P_2)$	OGLE				MACHO			
		$N_{\log P < 1}$	$N_{\log P > 1}$	$F$	$p(F)$	$N_{\log P < 1}$	$N_{\log P > 1}$	$F$	$p(F)$
0.4	1.5	585	49	8.867	0.000	1147	69	12.32	0.000
0.6	1.4	268	45	8.799	0.000	494	63	10.62	0.000
0.7	1.3	136	41	6.904	0.001	249	58	5.438	0.005
0.8	1.2	72	30	0.279	0.757	141	51	1.124	0.327
0.9	1.1	25	15	0.909	0.412	56	29	0.029	0.972
0.9	1.5	25	49	1.000	0.373	56	69	0.225	0.799

way of writing  $SSE$  as used in equation (5) is  $SSE = \text{Var}(\hat{a})(N - n)S_{XX}$ . Thus the  $F$ -test automatically builds into it the amount of precision in both slopes, which in turn is a function of the data used and the amount of variability about the regression line. Alternatively, one may think of the  $F$ -test as  $F = t^2$ , where  $t$  is equal to the difference in the two slopes divided by the standard error of the difference. Thus if we have another dataset with the same number of data points as ours but, for example, a larger (or smaller) scatter, then for this other dataset the slopes are less (or better) well defined with larger (or smaller) variance and the  $F$ -test statistic would generally decrease (or increase). In another situation, if we add (or remove) some data from our sample (assume the old and new samples have the same distribution), then generally the slopes will be more (or less) accurately estimated and the  $F$ -test statistic would increase (or decrease). We say “generally” because the data that is added or removed could have a large or small scatter which can, in some cases, offset the change in the amount of data. Thus the  $F$ -test statistic is sensitive to both the number of data points and the nature of the data.

We have verified this by using the LMC OGLE data, as used in KN, and formulating a  $F$ -test for a difference in the PL slope between short ( $P < 10$  days) and long ( $P > 10$  days) period Cepheids. However we also make a cut at periods  $P_1$  and  $P_2$  such that  $P_1 < 10 < P_2$ , thus decreasing the number of Cepheids available to define the slope on either side of 10 days. The results are presented in Table 2 for the OGLE  $V$ -band data used in KN, as well as the MACHO  $V$ -band data as described in Section 2 (the Sample C). This table shows that the  $F$ -value generally decrease, or the significance of the results generally goes down (with  $p$ -values increase), as the number of Cepheids on either side of the break point goes down<sup>2</sup>. It also depends, to some extent, on the nature of the data that was removed or added. We see that when we have less than about 30 Cepheids on the long period side, the significance is reduced and the  $F$ -test would indicate that the data are consistent with a single line. Similarly when we have less than about 25 Cepheids on the short period side. With such few Cepheids the slopes are not well defined with large errors. This gives us confidence that when we do get a significant result, that result has some meaning.

### 3.1 Choosing $P_o$

In the recent studies of the non-linear LMC PL relations (Tammann & Reindl 2002; Kanbur & Ngeow 2004; Sandage et al. 2004), the characteristic period  $P_o$  is chosen to be 10 days. Sandage et al. (2004) has discussed the significance of the discontinuity at 10 days, which will not be repeated here. However, the value of  $P_o$  can be found from equation (2) by making  $\log(P_o)$  another parameter in the least squares fit. To find  $P_o$ , we applied non-linear estimation procedures, as given in the SAS package, to the  $V$ -band data from Sample C. The result is  $\log(P_o) = 0.934$ , with the upper and lower 95% confidence limits of 1.089 and 0.778, respectively<sup>3</sup>. This result is fully consistent with the choice of  $P_o = 10$  days in Tammann & Reindl (2002), KN and Sandage et al. (2004). Other evidence to support the choice of  $P_o = 10$  days include:

(i) The amplitudes and the Fourier parameters from the light curves of the short period ( $P < 10$  days) Cepheids are radically different from their long period ( $P > 10$  days) counterparts (see Simon & Lee 1981; Simon & Moffett 1985; Andreasen 1988; Udalski et al. 1999a; Kanbur et al. 2002; Ngeow et al. 2003; Sandage et al. 2004, and references therein), where the plots of the amplitudes and Fourier parameters as a function of  $\log(P)$  show an abrupt change at 10 days. This is a well established fact in stellar pulsation. There is a well-known explanation for this change in the Fourier parameters at 10 days: the Hertzsprung progression (see, e.g. Simon & Lee 1981; Bono et al. 2000b, and references therein). We do not imply that this theoretical explanation for the change in Fourier parameters at a period of 10 days is responsible for possible changes in the Cepheid PL relations at the same period, merely that there is a precedent in the Cepheid literature for “something happening” at a period of 10 days.

(ii) In a separate paper (Ngeow & Kanbur, 2005 – in preparation), we study the multi-phase PC and PL relations for the OGLE LMC Cepheids. This is because the PL and PC relations at mean light (as, e.g., presented in Tammann & Reindl 2002; Kanbur & Ngeow 2004; Sandage et al. 2004) are the average of all other phases (Kanbur & Ngeow 2004; Kanbur et al. 2004). From this study, we found that both of the LMC PL and PC

<sup>2</sup> Note that the peak of the period distribution for both samples is located around  $\log(P) \sim 0.5$ , see Section 5.2.

<sup>3</sup> The result using Sample A is:  $\log(P_o) = 0.981$ , with the 95% confidence limits of 1.148 and 0.815.

relations provide compelling evidence of a strong non-linearity at period of 10 days at a pulsation phase of 0.82 (where phase zero corresponds to maximum light). A preliminary result from this study is presented in Figure 3, and the details will be given elsewhere.

(iii) The work of Tammann & Reindl (2002), KN and Sandage et al. (2004) strongly suggests that the LMC (optical) PC relation in  $(B - V)$  and/or  $(V - I)$  colour is also non-linear with a break at 10 days. This can be clearly see from figure 2 of Tammann & Reindl (2002), figure 2 of KN (which also include the PC relations at maximum and minimum light), figure 1 of Sandage et al. (2004) and also the left panels of Figure 5 in this paper.

(iv) The behavior of the amplitude-colour (AC) relations for the LMC Cepheids are different for the short ( $P < 10$  days) and long period Cepheids at maximum, mean and minimum light (KN). This is because the PC and AC relations are connected as described in KN.

(v) The residuals plots using one-line regression shows a trend at long period range ( $\log[P] \gtrsim 1.0$ ) for the  $V$ - and  $R$ -band PL relations (Figure 2 & right panels of Figure 4), as well as the PC relations (see figure 4 from Kanbur & Ngeow 2004).

(vi) The empirical LMC instability strip, as presented in Sandage et al. (2004), clearly show a break at a period of 10 days. The properties of the LMC instability strip, such as the ridge lines of the strip and the slope of constant-period lines, also different for the long ( $\log[P] > 1.0$ ) and short period Cepheids.

Hence, we follow the existing literature (Tammann et al. 2002; Kanbur & Ngeow 2004; Sandage et al. 2004) and adopt  $P_o = 10$  days in this paper. Note that due to the intrinsic dispersion of the PL relation (because of the finite width of the instability strip), the exact location of the  $P_o$  is difficult to accurately determine with statistical means from the data. We emphasize that the confirmation of  $P_o \simeq 10$  days has to be done with the stellar pulsation modeling, which is beyond the scope of this paper.

## 4 RESULTS

In Table 3 - 6, we present the fitted results for one-line, two-lines, piecewise continuous and quadratic regressions, respectively, for each of the three samples considered in this paper. It is worthwhile to point out that even though the dispersion values (given by  $\sqrt{\sum [m(i) - \hat{m}(i)]^2 / (N - n)}$ , which is not the same as the  $SSE$  that enters the chi-square test and the  $F$ -test) given in the last column of these tables are comparable to each other<sup>4</sup>, though with a slight improvement in Table 4 & 5 as compared to Table 3. This, however, does not mean that the four regression models describe the data equally well. The  $SSE$  is approximately the dispersion multiplied by the number of data points so differences between the models are more apparent. Further, a statistical test based purely on a comparison of the dispersion does not have the required power and is not a good

**Table 3.** The results of the one-line regressions in the form of  $m = a \log(P) + b$ , where  $\sigma$  is the dispersion of the regression.

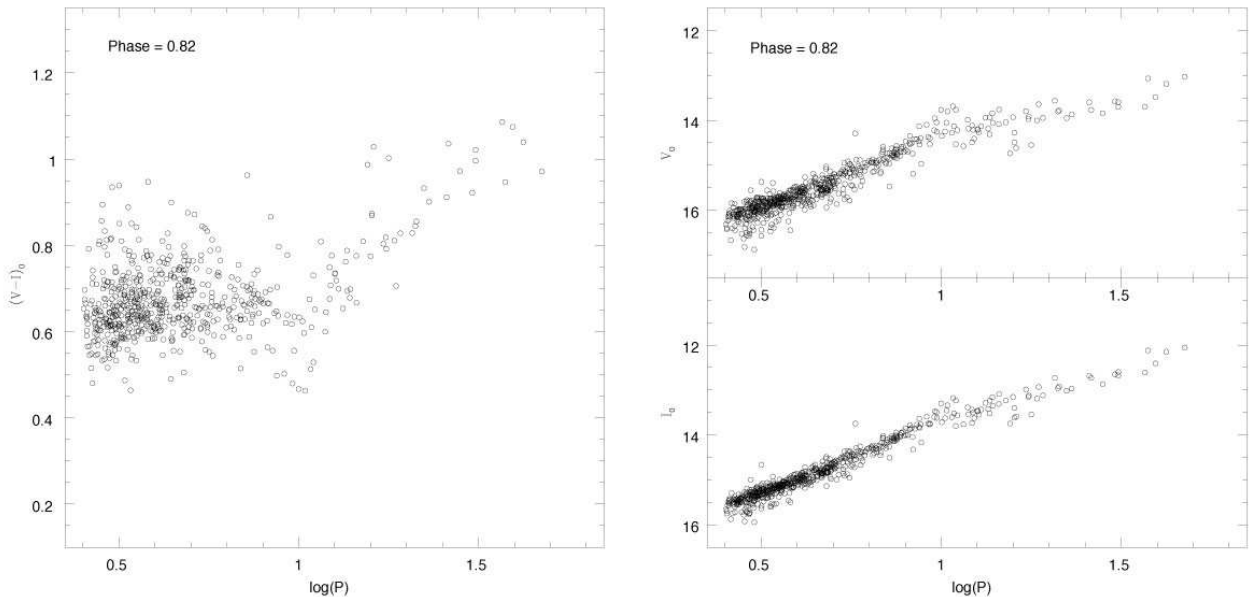
Bandpass	$a$	$b$	$\sigma$
Sample A, $N = 1330$			
$V \dots$	$-2.669 \pm 0.031$	$17.074 \pm 0.020$	0.285
$R \dots$	$-2.826 \pm 0.027$	$16.860 \pm 0.018$	0.244
$J \dots$	$-3.080 \pm 0.022$	$16.277 \pm 0.014$	0.130
$H \dots$	$-3.175 \pm 0.020$	$16.046 \pm 0.013$	0.114
$K \dots$	$-3.156 \pm 0.020$	$15.973 \pm 0.014$	0.129
Sample B, $N = 1322$			
$V \dots$	$-2.672 \pm 0.031$	$17.066 \pm 0.020$	0.257
$R \dots$	$-2.828 \pm 0.027$	$16.853 \pm 0.018$	0.220
$J \dots$	$-3.080 \pm 0.022$	$16.275 \pm 0.014$	0.124
$H \dots$	$-3.174 \pm 0.020$	$16.044 \pm 0.013$	0.112
$K \dots$	$-3.156 \pm 0.020$	$15.972 \pm 0.014$	0.128
Sample C, $N = 1216$			
$V \dots$	$-2.699 \pm 0.035$	$17.087 \pm 0.023$	0.259
$R \dots$	$-2.852 \pm 0.030$	$16.871 \pm 0.020$	0.222
$J \dots$	$-3.088 \pm 0.024$	$16.281 \pm 0.016$	0.124
$H \dots$	$-3.169 \pm 0.022$	$16.040 \pm 0.015$	0.110
$K \dots$	$-3.159 \pm 0.021$	$15.974 \pm 0.015$	0.124

statistical test to apply to our data. This is due to the following reasons: (a) the dispersion is dominated by the intrinsic dispersion ( $\sigma_{IS}$ ) because of the existence of the instability strip; (b) the majority of the Cepheids are in the short period range which dominates the dispersion; and (c) more importantly, the non-linear signature of the mean light LMC PL relation is not obvious as shown in, e.g., Figure 3, or with a large quadratic term ( $\gamma$ ) in equation (3). Therefore, it is necessary to apply the chi-square test and the  $F$ -test to detect such non-linear signature from our data.

The extinction corrected  $VRJHK$ -band PL relations are presented in the left panels of Figure 4 (from Sample C for illustrative purposes, together with the fitted two-lines regressions. The PL relations from the other two samples, as well as other forms of regressions, are similar to Figure 4). From Figure 4, the  $V$ - and  $R$ -band PL relations imply that there is a break or sharp non-linearity at/around 10 days, marginally in the  $J$ -band PL relation, but not obvious in the  $H$ - and  $K$ -band PL relations. The right panels of Figure 4 compare the residuals of the fit as a function of period using one-line and two-line regressions (again the residual plots from the two other forms of regressions are very similar to these plots). From the residual plots, both of the  $V$ - and  $R$ -band residuals show a trend at  $\log(P) > 1.0$  when using the one-line regression. These trends are removed if two-line regressions are used to fit the data. These findings also support a break in the PL relation at a period of 10 days. Similarly, the  $J$ -band residuals show a marginal trend, which is not obvious in the  $H$ - and  $K$ -band.

Both of the chi-square test and the  $F$ -test were then applied to the samples. The results from the chi-square test and the  $F$ -test are summarized in Table 7 & 8, respectively. These results can be compared with Table 1 to decide if the null hypothesis can be rejected. If the  $\chi^2$  and/or the  $F$  values given in Table 7 & 8 are smaller than the values corresponding to  $p = 0.05$ , then the underlying PL relation is linear. In contrast, if the  $\chi^2$  and/or the

<sup>4</sup> We thank the anonymous referee to point this out.



**Figure 3.** LMC extinction corrected PC (left panel) and PL (right panel) relations at phase of 0.82 (where phase zero = maximum light), obtained from using the OGLE photometric data. A discontinuity at/near 10days is clearly visible.

$F$  values in these tables are larger than the values corresponding to  $p = 0.01$ , then the underlying PL relation is non-linear. For the  $\chi^2$  and/or the  $F$  values that fall in-between the values corresponding to  $p = 0.05$  and  $p = 0.01$ , we classify the PL relation to be marginally linear. In short, from Table 7 & 8 we found that:

(i) For  $H_A$  = two-line and piecewise continuous regressions, both of the chi-square test and the  $F$ -test show that the  $VRJ$  &  $H$ -band data (in all three samples) are more consistent with the two-line regression, and the null hypothesis with one-line regression can be rejected with very high confidence level ( $> 99.9\%$ ). In contrast, the  $K$ -band data (in all three samples) is consistent with the one-line regression from the chi-square test. The  $F$ -test results for the  $K$ -band data also show that the  $K$ -band PL relation is marginally linear<sup>5</sup>.

(ii) For  $H_A$  = quadratic regression, the results are quite different to the two other types of regression. The linear/non-linear behavior is quite different in  $VRJ$ - and  $H$ -band data across the three samples. From the chi-square test, the  $VRJ$ -band PL relations are linear in Sample A, then become marginally linear in Sample B and non-linear in Sample C. The  $H$ -band PL relation is marginally linear in Sample A but non-linear in both of the Sample B & C from the chi-square test. In terms of the  $F$ -test, Sample A shows that the  $V$ - &  $R$ -band PL relations are linear, marginally linear for the  $J$ -band PL relation and non-linear for the  $H$ -band PL relation.

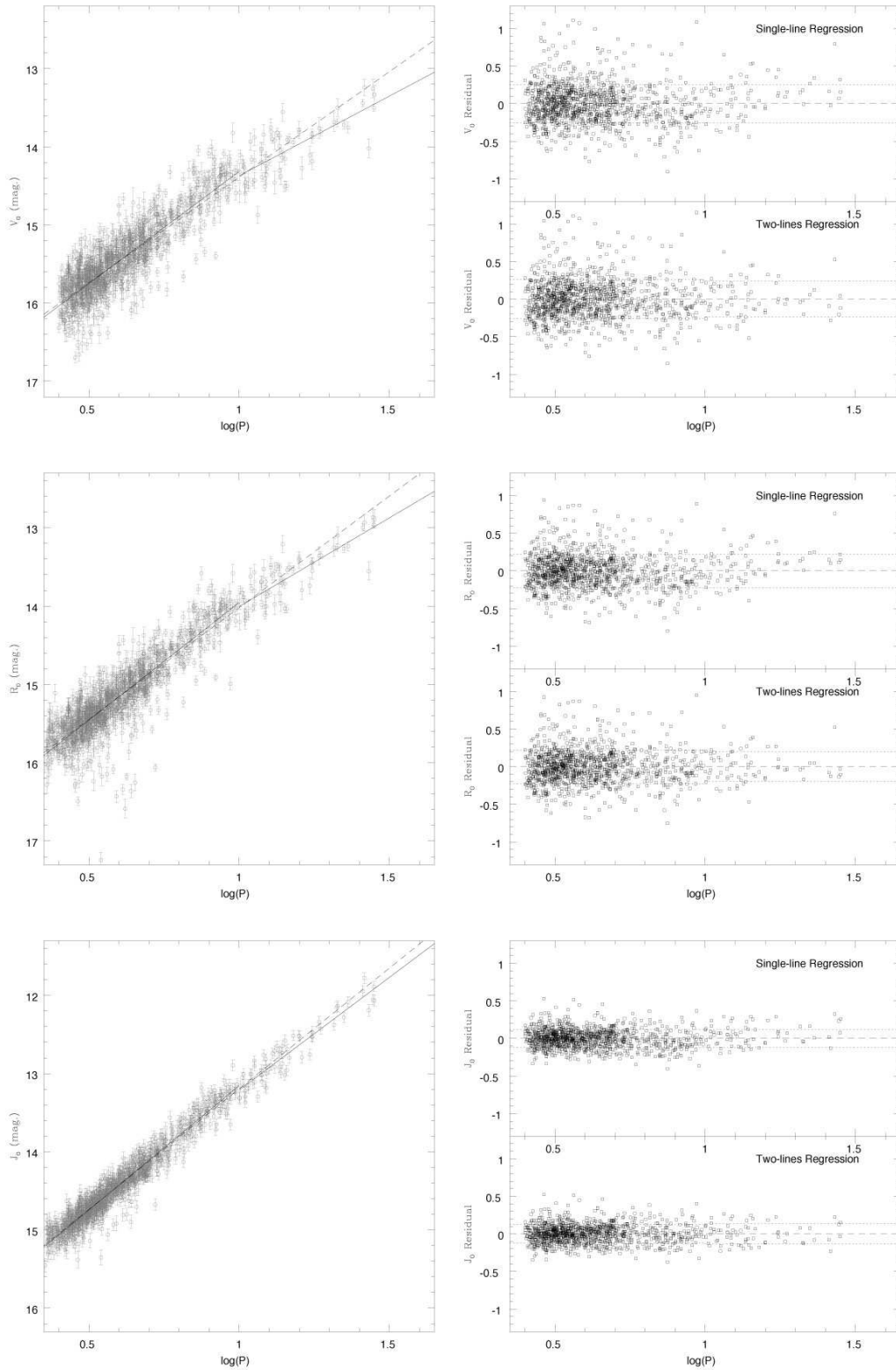
<sup>5</sup> Note that for the  $K$ -band results, the  $F$ -values for  $H_A$  = piecewise continuous regression are larger than the  $F$ -values from  $H_A$  = two-line regression. Hence the  $K$ -band PL relation will be more consistent with the piecewise continuous regression if  $p = 0.01$  is chosen and vice versa.

In Sample B, the  $V$ -band PL relation is still linear, but the  $R$ -band PL relation becomes marginally linear, and both of the  $J$ - and  $H$ -band PL relations are non-linear. The  $VRJ$ - and  $H$ -band PL relations become non-linear in Sample C. In contrast, the  $K$ -band PL relations are linear in all three samples from both of the chi-square test and the  $F$ -test.

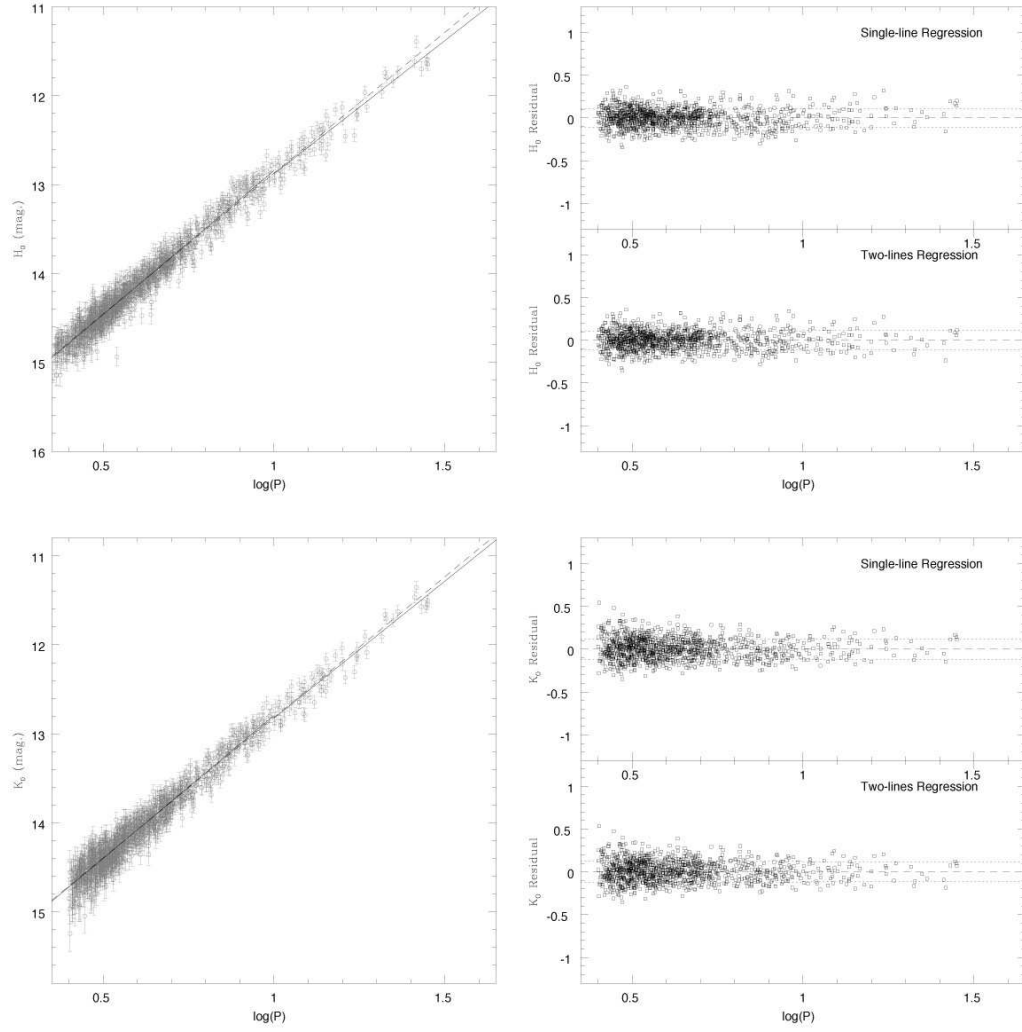
In terms of the sample selection, the results from Sample A are not a good choice as the final adopted results. This is because Sample A suffers from obvious outliers as shown in Figure 1. Further, these outliers have made the quadratic fit for the  $VRJ$ -band PL relations to be more linear (as the quadratic term in equation [3],  $\gamma$ , is consistent with zero, see Table 6), in contrast to the results from the two-line or piecewise continuous regressions. Hence the results from Sample A is not considered further. Both of the Sample B and Sample C are freed from the obvious outliers, with the difference that Sample B includes Cepheids with  $\log(P) < 0.4$  but Sample C excludes these very short period Cepheids. The statistical tests with two-line and piecewise continuous regressions show consistent results for both samples, but the results with quadratic regression are slightly different as mentioned previously. This implies that neglecting Cepheids with  $\log(P) < 0.4$  does not influence our results. We favor the results from Sample C because there are some physical reasons to exclude the Cepheids with  $\log(P) < 0.4$ , as discussed in Section 2 (the contamination from overtone Cepheids and the possible changes of slope due to evolutionary effect).

Among the three full non-linear models, the true form of the non-linearity is difficult to distinguished statistically from the data. This has to be done together with the theoretical calculations, which is beyond the scope of this paper. However, we favor both two-line and piece-





**Figure 4.** *Left (a):* The extinction corrected  $VRJHK$ -band PL relations for Sample C. The dashed and solid lines are the fitted one-line and two-line regressions, respectively. The error bars are the quadrature sum of the photometric errors in mean magnitudes ( $\sigma_\lambda$ ), the errors from the random phase corrections that are only applicable to  $JHK$ -bands ( $\sigma_{rp}$ ) and the errors from the extinction ( $\sigma_{A_\lambda}$ ). *Right (b):* The residuals from using one-line (top panel) and two-line (bottom panel) regression fits. The dashed line indicates the zero residual. The dotted lines are the  $1\sigma$  dispersions.

Figure 4. *continued.***Table 4.** The results of the two-line regressions in the form of  $m = a_{(S,L)} \log(P) + b_{(S,L)}$ , where  $\sigma_{(S,L)}$  is the dispersion of the regression. The subscripts  $S,L$  are referred to the short ( $P < 10$  days) and long period Cepheids, respectively.

Bandpass	$a_S$	$b_S$	$\sigma_S$	$a_L$	$b_L$	$\sigma_L$
Sample A, $N_S = 1261$ , $N_L = 69$						
$V \dots$	$-2.771 \pm 0.041$	$17.130 \pm 0.025$	0.286	$-2.039 \pm 0.222$	$16.409 \pm 0.258$	0.238
$R \dots$	$-2.919 \pm 0.036$	$16.911 \pm 0.022$	0.245	$-2.263 \pm 0.193$	$16.267 \pm 0.224$	0.200
$J \dots$	$-3.143 \pm 0.029$	$16.312 \pm 0.018$	0.129	$-2.884 \pm 0.151$	$16.093 \pm 0.175$	0.132
$H \dots$	$-3.234 \pm 0.027$	$16.078 \pm 0.017$	0.113	$-3.003 \pm 0.134$	$15.883 \pm 0.155$	0.117
$K \dots$	$-3.193 \pm 0.028$	$15.994 \pm 0.017$	0.129	$-3.080 \pm 0.125$	$15.904 \pm 0.145$	0.113
Sample B, $N_S = 1253$ , $N_L = 69$						
$V \dots$	$-2.787 \pm 0.041$	$17.129 \pm 0.025$	0.256	$-2.039 \pm 0.222$	$16.409 \pm 0.258$	0.238
$R \dots$	$-2.933 \pm 0.036$	$16.910 \pm 0.022$	0.220	$-2.263 \pm 0.193$	$16.267 \pm 0.224$	0.200
$J \dots$	$-3.148 \pm 0.029$	$16.312 \pm 0.018$	0.123	$-2.884 \pm 0.151$	$16.093 \pm 0.175$	0.132
$H \dots$	$-3.235 \pm 0.027$	$16.078 \pm 0.017$	0.110	$-3.003 \pm 0.134$	$15.883 \pm 0.155$	0.117
$K \dots$	$-3.195 \pm 0.028$	$15.994 \pm 0.017$	0.129	$-3.080 \pm 0.125$	$15.904 \pm 0.145$	0.113
Sample C, $N_S = 1147$ , $N_L = 69$						
$V \dots$	$-2.867 \pm 0.048$	$17.184 \pm 0.030$	0.257	$-2.039 \pm 0.222$	$16.409 \pm 0.258$	0.238
$R \dots$	$-3.005 \pm 0.042$	$16.959 \pm 0.026$	0.221	$-2.263 \pm 0.193$	$16.267 \pm 0.224$	0.200
$J \dots$	$-3.179 \pm 0.033$	$16.333 \pm 0.021$	0.122	$-2.884 \pm 0.151$	$16.093 \pm 0.175$	0.132
$H \dots$	$-3.239 \pm 0.030$	$16.081 \pm 0.019$	0.109	$-3.003 \pm 0.134$	$15.883 \pm 0.155$	0.117
$K \dots$	$-3.207 \pm 0.030$	$16.002 \pm 0.019$	0.125	$-3.080 \pm 0.125$	$15.904 \pm 0.145$	0.113

**Table 5.** The results of the piecewise continuous regressions in the form of  $m = c + d \times \min[\log(P), 1.0] + e \times \max[0, \log(P) - 1.0]$ , where  $\sigma$  is the dispersion of the regression.

Bandpass	$c$	$d$	$e$	$\sigma$
Sample A, $N = 1330$				
$V \dots$	$17.128 \pm 0.024$	$-2.767 \pm 0.038$	$-2.000 \pm 0.155$	0.284
$R \dots$	$16.908 \pm 0.021$	$-2.915 \pm 0.033$	$-2.219 \pm 0.135$	0.243
$J \dots$	$16.305 \pm 0.017$	$-3.130 \pm 0.027$	$-2.751 \pm 0.106$	0.129
$H \dots$	$16.072 \pm 0.016$	$-3.221 \pm 0.025$	$-2.889 \pm 0.094$	0.113
$K \dots$	$15.989 \pm 0.016$	$-3.184 \pm 0.025$	$-3.004 \pm 0.089$	0.129
Sample B, $N = 1322$				
$V \dots$	$17.125 \pm 0.024$	$-2.778 \pm 0.038$	$-1.944 \pm 0.155$	0.255
$R \dots$	$16.906 \pm 0.021$	$-2.925 \pm 0.033$	$-2.170 \pm 0.135$	0.219
$J \dots$	$16.304 \pm 0.017$	$-3.133 \pm 0.027$	$-2.735 \pm 0.106$	0.124
$H \dots$	$16.071 \pm 0.016$	$-3.222 \pm 0.025$	$-2.880 \pm 0.094$	0.111
$K \dots$	$15.989 \pm 0.016$	$-3.185 \pm 0.025$	$-2.999 \pm 0.089$	0.128
Sample C, $N = 1216$				
$V \dots$	$17.172 \pm 0.028$	$-2.846 \pm 0.044$	$-1.863 \pm 0.157$	0.256
$R \dots$	$16.948 \pm 0.024$	$-2.985 \pm 0.038$	$-2.098 \pm 0.137$	0.219
$J \dots$	$16.321 \pm 0.019$	$-3.157 \pm 0.030$	$-2.708 \pm 0.107$	0.123
$H \dots$	$16.072 \pm 0.018$	$-3.223 \pm 0.028$	$-2.879 \pm 0.095$	0.109
$K \dots$	$15.995 \pm 0.018$	$-3.194 \pm 0.028$	$-2.988 \pm 0.090$	0.124

**Table 6.** The results of the quadratic regressions in the form of  $m = \alpha + \beta \log(P) + \gamma [\log(P)]^2$ , where  $\sigma$  is the dispersion of the regression.

Bandpass	$\alpha$	$\beta$	$\gamma$	$\sigma$
Sample A, $N = 1330$				
$V \dots$	$17.126 \pm 0.050$	$-2.829 \pm 0.144$	$0.110 \pm 0.097$	0.285
$R \dots$	$16.914 \pm 0.044$	$-2.993 \pm 0.125$	$0.115 \pm 0.084$	0.244
$J \dots$	$16.336 \pm 0.036$	$-3.259 \pm 0.102$	$0.123 \pm 0.068$	0.130
$H \dots$	$16.125 \pm 0.035$	$-3.412 \pm 0.098$	$0.159 \pm 0.065$	0.114
$K \dots$	$15.996 \pm 0.037$	$-3.222 \pm 0.102$	$0.043 \pm 0.066$	0.129
Sample B, $N = 1322$				
$V \dots$	$17.160 \pm 0.050$	$-2.961 \pm 0.144$	$0.200 \pm 0.097$	0.257
$R \dots$	$16.944 \pm 0.044$	$-3.108 \pm 0.125$	$0.193 \pm 0.084$	0.220
$J \dots$	$16.346 \pm 0.036$	$-3.296 \pm 0.103$	$0.148 \pm 0.068$	0.124
$H \dots$	$16.131 \pm 0.035$	$-3.434 \pm 0.098$	$0.174 \pm 0.065$	0.111
$K \dots$	$16.000 \pm 0.037$	$-3.236 \pm 0.102$	$0.052 \pm 0.066$	0.128
Sample C, $N = 1216$				
$V \dots$	$17.345 \pm 0.072$	$-3.435 \pm 0.197$	$0.477 \pm 0.126$	0.258
$R \dots$	$17.114 \pm 0.063$	$-3.545 \pm 0.171$	$0.449 \pm 0.109$	0.221
$J \dots$	$16.427 \pm 0.050$	$-3.502 \pm 0.136$	$0.267 \pm 0.086$	0.123
$H \dots$	$16.159 \pm 0.046$	$-3.505 \pm 0.124$	$0.215 \pm 0.078$	0.109
$K \dots$	$16.031 \pm 0.046$	$-3.317 \pm 0.123$	$0.100 \pm 0.077$	0.124

wise continuous regression model over the polynomial fit because of the following reasons: (a) the residual plots in Figure 4 suggest that the two-line/piecewise continuous regression is more applicable; (b) the (optical) PC relations are more consistent with two-line or piecewise continuous regression (see left panels of Figure 5 below and the PC relation presented in Tammann et al. 2002; Kanbur & Ngeow 2004; Sandage et al. 2004); (c) a simple period-mean density relation predicts a PL relation linear in  $\log(P)$  but not in  $[\log(P)]^2$ ; and (d) Figure 3 shows that the two-line or the piecewise continuous regressions are more appropriate to describe the data. Nevertheless it is clear this question is rather moot since a two-line or a piecewise continuous model with two separate slopes for short and long period Cepheids approximates a quadratic

PL relation. For the (4-parameter) two-line regression model vs. the piecewise continuous regression model, it is true that either model may be appropriate. Again this has to wait for the future theoretical calculations to answer this problem.

In short, from the above discussion, the main results we found from this study are:

*Based on the data used in this study, the LMC V- & R-band relations are non-linear around a period of 10 days. Assuming that veracity of the random phase corrections used for the JHK data, we also find the J- & H-band PL relations are also non-linear across this same period but the K-band PL relation is (marginally) linear.*

## 5 CONCLUSION AND DISCUSSION

By applying the statistical tests to the *VRJHK*-band PL relations obtained from the MACHO and 2MASS data, we find that the *VRJH*-band PL relations show strong evidence that they are non-linear while the *K*-band PL relation is (marginally) consistent with a linear PL relation. The non-linearity of the extinction corrected *V*-band PL relations as seen from two independent datasets, one from the OGLE data (Tammann et al. 2002; Kanbur & Ngeow 2004; Sandage et al. 2004) and another from this study, strongly indicates that this non-linearity is real and not due to the artifacts of, for example, the extinction error and/or the photometric reductions. Note that Persson et al. (2004) suggested the NIR PL relations (based on 88 LMC Cepheids) are linear. Their results may well be influenced by the small number of short period Cepheids (18) in their sample. Recall that Table 2 indicates how small numbers of short and/or long period Cepheids can lead to an *F*-test indicating consistency with one-line regression. We can apply the *F*-test to the 88 Cepheids (after the extinction correction) given in Persson et al. (2004), including the 18 short period Cepheids. The *F*-test results are:  $F_J = 2.825$  ( $p = 0.065$ ),  $F_H = 2.356$  ( $p = 0.101$ ) and  $F_K = 2.215$  ( $p = 0.116$ ). These results indicate that the corresponding PL relations are linear, as expected (from Table 2), and as reported in Persson et al. (2004), though the result for the *J*-band is marginal. The reason for these non-significant results is because of the small number of short period Cepheids (however, see Section 5.2), leading to a short period slope that has a large error. However, we also see that our *F*-test results for this dataset has decreasing significance as the wavelength increases. This is exactly what we see in our results for the MACHO/2MASS *JHK* data.

The reason that the PL relation is non-linear is because the PC relation for the LMC Cepheids is also non-linear (Tammann et al. 2002; Kanbur & Ngeow 2004). This can be clearly seen from upper panels of Figure 5 for extinction corrected ( $V - R$ ) & ( $V - K$ ) PC relations. Since both of the PL and PC relations obey the more general PLC relations, the morphology of the PC relation will affect the morphology of the PL relation (Madore & Freedman 1991). Recall that the luminosity variation for the Cepheid variables in visual bands is dominated by the temperature variation (Cox 1980). However, as the waveband moves toward the infrared, the radius variation begin to dominate the variation in the infrared PL relations (e.g., Cox 1980; Madore 1985). Since there is no obvious reason that the period-radius relation for LMC Cepheids should be non-linear (see, e.g., Bono et al. 1998; Gieren et al. 1998, 1999), one would expect that the PL relation is linear toward the infrared. From this study, we see that the PL relation is non-linear at *VRJ*- & *H*-band, and marginally linear at *K*-band. This suggests that the radius variation may begins to dominate at *K*-band and longer wavelengths. This evidence can also be seen from the lower-right panel of Figure 5, where the extinction corrected ( $J - K$ ) PC relation becomes linear. These findings also imply that the non-linearity of the PL relations is due to the temperature (or colour) variations

**Table 7.** The results of the chi-square test, with  $H_O$  = one-line regression and  $\log(P_o) = 1.0$ .

$H_A =$ Bandpass	two-lines $\chi^2$	piecewise $\chi^2$	quadratic $\chi^2$
Sample A, $N = 1330$			
<i>V</i> ...	19.47	19.41	1.282
<i>R</i> ...	21.26	21.16	1.869
<i>J</i> ...	11.69	10.14	3.222
<i>H</i> ...	11.14	9.687	6.063
<i>K</i> ...	3.870	3.127	0.426
Sample B, $N = 1322$			
<i>V</i> ...	23.35	22.99	4.235
<i>R</i> ...	25.34	24.89	5.235
<i>J</i> ...	13.06	11.14	4.652
<i>H</i> ...	11.91	10.23	7.276
<i>K</i> ...	4.166	3.321	0.631
Sample C, $N = 1216$			
<i>V</i> ...	30.85	29.60	14.38
<i>R</i> ...	33.40	31.94	16.86
<i>J</i> ...	16.02	13.26	9.553
<i>H</i> ...	11.52	9.778	7.527
<i>K</i> ...	4.981	3.877	1.698

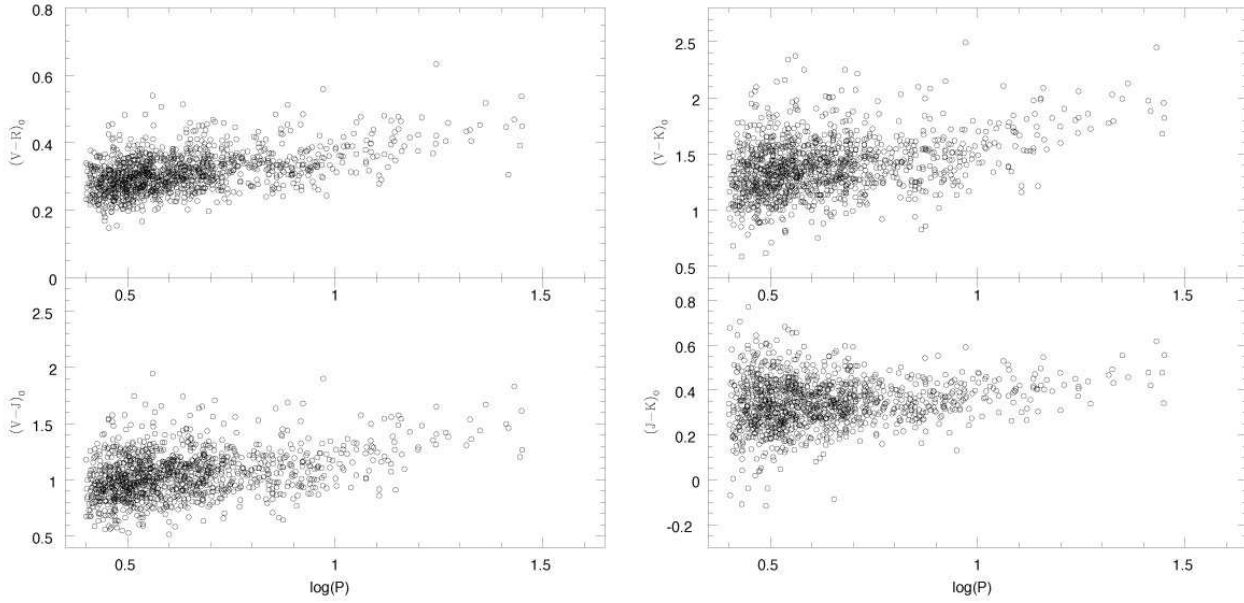
as a function of period. The detailed discussion and investigation of the physical reason behind the non-linear PL and PC relation is beyond the scope of this paper, and it will be addressed in a future paper.

Some attempts at using the long period LMC PL relation (from two-line regression) in distance scale studies can be found in Kanbur et al. (2003), Leonard et al. (2003), Thim et al. (2003), Thim et al. (2004) and Riess et al. (2005) without any statistical justification. We have shown in Section 3 & 4 that there is strong statistical evidence that the (optical) LMC PL relation is non-linear, hence the use of long period PL relation is justified. Ngeow & Kanbur (2005) has discussed the effect of using the broken LMC PL relations (two-line regression) in distance scale applications: they found that this effect is minimal with a systematic error of 0.03 mag. or less for the derived distance modulus. Therefore the finding of a non-linear LMC PL relation is more important in terms of the studies of stellar pulsation and evolution to understand the physical reasons that cause the non-linearity than in distance scale applications. However, in the era of precision cosmology, it is important to use the latest understanding of Cepheid PL relations.

We now discuss some relevant issues that may affect our results in the following subsections.

### 5.1 Issue of outliers

Could outliers be one cause for our results? – KN tested this extensively for both PL and PC relations and concluded that outliers cannot be responsible for the break. Figure 3, using OGLE data and depicting both PC and PL relations at pulsation phase close to 0.82, is remarkably clear of outliers: the break or non-linearity at a period of 10 days is compelling. The fact that our analysis of MACHO/2MASS data produce results consistent with our previous analysis of the OGLE data set suggests that outliers cannot be responsible for the results presented



**Figure 5.** The selected PC relations after extinction correction, constructed with the Sample C data as mentioned in Section 2.

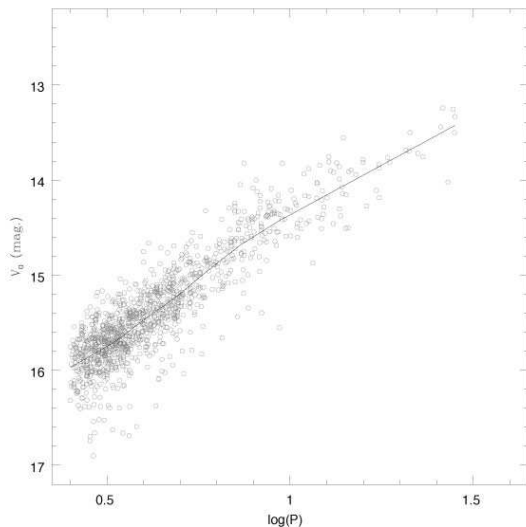
**Table 8.** The results of the  $F$ -test, with  $H_O$  =one-line regression and  $\log(P_o) = 1.0$ .

$H_A =$ Bandpass	two-lines $F$	piecewise $F$	quadratic $F$
Sample A, $N = 1330$			
$V \dots$	6.367	12.71	0.832
$R \dots$	7.178	14.30	1.251
$J \dots$	8.988	15.58	4.908
$H \dots$	9.539	16.57	10.32
$K \dots$	2.923	4.723	0.642
Sample B, $N = 1322$			
$V \dots$	9.428	18.58	3.383
$R \dots$	10.53	20.70	4.301
$J \dots$	10.91	18.57	7.692
$H \dots$	10.68	18.33	12.98
$K \dots$	3.182	5.072	0.960
Sample C, $N = 1216$			
$V \dots$	12.32	23.64	11.37
$R \dots$	13.76	26.31	13.75
$J \dots$	13.37	22.07	15.83
$H \dots$	10.30	17.46	13.40
$K \dots$	3.788	5.891	2.573

in this paper. Nevertheless, we adopt a robust regression method which limits the influence of outliers by using Tukey's Bi-weight function as, e.g., given in Press et al. (1992) to the  $V$ -band data with two-line vs. one-line regression. The general idea is to suppress the influence of points with large deviation from the expected model. The  $F$ -test results are:  $F_V = 5.869$  ( $p = 0.003$ ) for Sample A,  $F_V = 5.924$  ( $p = 0.003$ ) for Sample B and  $F_V = 9.196$  ( $p = 0.000$ ) for Sample C, respectively. Minimizing the sum of the absolute deviations rather than the sum of the squares also yields broadly similar results. Another approach is the RANSAC algorithm (Fisher & Bolles 1981;

Storkey et al. 2004) used to detect linear structure in the presence of many outliers. This algorithm tries to find out how much support a given model has among the data. Equation (1) of Storkey et al. (2004) shows how many times we have to repeat the algorithm for a given probability of failure. Given that we have over  $\sim 1300$  data points this algorithm only makes sense if we have over  $\sim 200$  outliers and we do not believe this to be the case. We also tried the sigma-clipping algorithm (Udalski et al. 1999a) to remove more outliers, and again we found that our results remain unchanged.

All the above methods are parametric and make specific assumptions about the form of the regression surface. The LOESS (Local Linear Smoothing) procedure in SAS implements a non-parametric method for implementing a non-parametric regression surface (Cleveland 1979). If we have  $y(i) = g[x(i)] + \epsilon(i)$ , where  $x, y$  are the predictors and independent variables respectively, LOESS works by approximating the value,  $g$ , of the regression function locally, say at  $x_0$ , by the value of a function in some specified parametric class. This is done by fitting a regression surface within a chosen neighborhood of  $x_0$ . The fraction of the data points contained in this neighborhood is the smoothing parameter. This procedure is suitable when there are outliers and robust techniques are required. Figure 6 portrays our results using this method when applied to the extinction corrected MACHO  $V$ -band data in Sample C. The solid line in this figure is the non-parametric fit and the points are the original data. The non-linearity of the fit between  $0.8 < \log(P) < 1.1$  is self-evident, and these results are insensitive to the smoothing parameter used. What we take from this is that our results are, to a large extent, independent of fitting method and robust to outliers.



**Figure 6.** A robust non-parametric fit (solid line) between  $\log(P)$  and the extinction corrected  $V$ -band MACHO data.

## 5.2 Issue of long period Cepheids

Could the small number/incompleteness at the long period end be responsible to our results? – If this is the case, then these effects should act in the same way at all phases. Again Figure 3 displays a preliminary result: the reddening corrected PC and PL relations for OGLE data at the pulsation phase of 0.82 in the left and right panels, respectively. If the true underlying PC relation at this phase is indeed linear, then the combination of selection effects/incompleteness/period cuts acts so as to make the longer period Cepheids appear cooler than they really are. From observations, longer period Cepheids obey a flat PC relation (Code 1947; Simon et al. 1993; Kanbur & Ngeow 2004; Kanbur et al. 2004; Kanbur & Ngeow 2005). Figure 2 of KN provides evidence for this for LMC Cepheids with the OGLE data. We have also evaluated the  $(V - R)$  colour at maximum light for the MACHO LMC Cepheid sample and also find a flat PC relation for these stars at maximum light (Kanbur et al 2005 – in preparation). However, if a combination of small number/incompleteness at the long period end/selection effects is responsible for our results, then the true distribution at maximum light must be such that longer period Cepheids get hotter or bluer as the period increases<sup>6</sup>. There is very credible physics that predicts the true PC relation at maximum light should be flat (Simon et al. 1993; Kanbur et al. 2004). Thus it is our contention that a combination of selection/incompleteness/period cuts cannot be responsible for our results.

The small number of long period Cepheids is expected as Alcock et al. (1999) and Udalski et al. (1999b) showed that the period distribution for the LMC Cepheids peaked at  $\log(P) \sim 0.5$  with a long tail ex-

tended out to the long period end. This is because in general Cepheids with longer period have higher mass. It is well known that due to the initial mass function ( $dN/dM \propto M^{-\alpha}$ , where  $M$  is the mass of the stars and  $\alpha > 0$ ), the number of high mass stars is expected to be less than the low mass stars (see, e.g., Alcock et al. 1999). In addition, the crossing time (across the instability strip) for the high mass stars is shorter than their low mass counterparts (Bono et al. 2000a). These two effects reduce the number of long period Cepheids in comparison to the number of short period Cepheids. To show our results do not suffer from the missing long period Cepheids with  $\log(P) > 1.5$ , we add extinction corrected  $V$ -band data for the 18 longest period (i.e.  $\log[P] > 1.5$ ) Cepheids from Sandage et al. (2004) to our sample. We then apply the  $F$ -test to the combined sample and test whether the data are more consistent with a one-line or two-line regression. The resulting  $F$ -value is 8.734, which implies that the LMC  $V$ -band PL relation is still non-linear with the additional data. We also tried to include all 97 Cepheids (with  $0.4 < \log[P] < 2.2$ ) from Sandage et al. (2004) to our sample, and the  $F$ -test result still supports the non-linear PL relation.

A more rigorous test, as suggested by an anonymous referee, is to increase the dataset by adding long period Cepheids from a homogeneous dataset that does not belong to the MACHO sample. Such data exist in the literature, for example from Sebo et al. (2002) which we will use in our test. We first match the Cepheid coordinates and periods in our MACHO data and Sebo et al. (2002) sample to identify the common Cepheids in both samples. We find 203 such stars. For these common Cepheids, we calculate the average difference,  $\overline{\Delta}$ , between the  $V$  and  $R$  mean magnitudes in these two samples, in the sense that  $\Delta_V = V_{\text{MACHO}} - V_{\text{Sebo}}$  and  $\Delta_R = R_{\text{MACHO}} - R_{\text{Sebo}}$ . After removing some obvious outliers ( $\overline{\Delta} > 1$ ), we obtain:  $\overline{\Delta}_V = 0.041$  (with std = 0.121) and  $\overline{\Delta}_R = 0.039$  (with std = 0.120). This step is to ensure the Sebo et al. (2002) photometric system is the same as MACHO photometric system by adding  $\overline{\Delta}$  to the Sebo et al. (2002) mean magnitudes. In addition, we also found that there is no period dependency in this transformation:  $\Delta_V = 0.037(\pm 0.034) \log(P) + 0.013(\pm 0.027)$ ,  $\Delta_R = -0.028(\pm 0.034) \log(P) + 0.060(\pm 0.027)$ . If there are any dependencies with period in such a transformation, it could be that this has influenced our results. We see clearly that this is not the case. Similarly, the average difference in  $JHK$  mean magnitudes between the Persson et al. (2004) sample and the MACHO/2MASS sample is negligible, and more importantly, independent of periods (see Section 2 for details).

These photometric transformations (albeit very close to zero) can now be used to transform the Sebo et al. (2002) data to the same photometric system as the MACHO  $V$ - and  $R$ -band data used in this study, and to transform the MACHO/2MASS  $JHK$  data to the same system as the Persson et al. (2004) data. This procedure provides a consistent way of increasing our long period sample by identifying those stars not in the MACHO/2MASS sample but which are in either the Sebo et al. (2002) or Persson et al. (2004) samples. We now describe our results when this is carried out.

<sup>6</sup> Note that Cepheids can generally get cooler as periods increase, due to the morphology of the instability strip.

The long period ( $\log[P] > 1.0$ ) Cepheids in the Sebo et al. (2002) sample that are not in the MACHO sample are readily identified. We found 49 of them<sup>7</sup> and they are listed in Table 9. In this table, we give the original  $V$  and  $R$  mean magnitudes from Sebo et al. (2002), as well as the number of observed data points ( $m_{V,R}$ ) for each Cepheids in either bands. The periods are adopted from either Persson et al. (2004) or the OGLE database (Udalski et al. 1999b). The table also lists the  $JHK$  mean magnitudes and the  $E(B-V)$  values for these Cepheids that are compiled from Persson et al. (2004) and (Udalski et al. 1999b), or we set  $E(B-V) = 0.10$  if no entries are given in Persson et al. (2004). Since the Sebo et al. (2002) database lacks Cepheids between  $\log(P) \sim 1.55$  and  $\log(P) \sim 1.95$ , we include 8 Cepheids in this period range (with the available  $JHK$  mean magnitudes) from Persson et al. (2004) in Table 9 for completeness. We did not attempt to add the  $V$  and  $R$  mean magnitude for these 8 Cepheids from the literature, as in Sandage et al. (2004), in order to keep our dataset as homogeneous and consistent as possible.

The data in Table 9 are then added to our MACHO/2MASS sample, after the appropriate photometric transformation. The  $F$ -tests are applied to this extended sample for  $H_A$  = two-lines regression (with a break at 10 days) using the following selection criteria:

- (i) Include all available long period Cepheids in Table 9, appropriate to the bandpass, to our MACHO/2MASS sample.
- (ii) Eliminate Cepheids without  $JHK$  mean magnitudes since not all of the Cepheids in Table 9 have the  $JHK$  means.
- (iii) Eliminate Cepheids with  $m_{V,R} \leq 10$  as the mean magnitudes derived from these sparsely sampled light curves may not be accurate.
- (iv) Combine step (ii) & (iii) above (item 5 in Table 10) in order to have consistent number of Cepheids in all five bands.

The results from these different selection criteria are presented in Table 10. Note that Persson et al. (2004) did not use Cepheids with  $\log(P) > 2.0$ . These Cepheids are eliminated in step (iv) above (item 5 in Table 10). The linearity of the PL relation in each band again can be tested by comparing the  $F$ -values in this table with Table 1. As can be seen from the table, including the additional long period Cepheids does not alter the main conclusion we have: the  $VRJH$ -band PL relation is non-linear and the  $K$ -band PL relation is (marginally) linear. Since there are no measurement errors given in Sebo et al. (2002) for the  $V$  and  $R$  mean magnitudes, the results in Table 10 are derived with unweighted regression (i.e.  $w = 1$ ) in all bands. The  $F$ -values from using the weighted regressions to all bands, with assumed some values for the errors in the  $V$ - and  $R$ -band means (and the  $E[B-V]$ , the errors for  $JHK$  means are given in Persson et al. 2004), are very similar to Table 10. One exception is the  $F$ -value for the

$K$ -band when using all available long period Cepheids, the  $F$ -value goes down from 2.942 to 1.532, which is fully consistent with a one-line regression. In this table, except for the  $K$ -band, the lowest significant  $F$ -value is 7.872. With the degrees of freedom appropriate for the data in our study, this has a  $p$ -value less than  $10^{-3}$  (recall that  $p < 10^{-3}$  for  $F > 7$ ). This means that assuming the null hypothesis of a single regression line being sufficient. The chance of getting a value of the  $F$  statistic as high or greater than that actually observed (7.872) is less than  $10^{-3}$ .

### 5.3 Issue of reddening/extinction

Could reddening and extinction errors produce our results? – Figure 5 displays extinction corrected PC relations, at mean light, obtained from the MACHO+2MASS data for a variety of colours:  $(V-R)_0$ ,  $(V-J)_0$ ,  $(V-K)_0$  &  $(J-K)_0$ . This figure and the PC relations presented in KN and Sandage et al. (2004) imply that in order to make a linear PC relation, extinction errors would have to be of the order of  $\sim 0.05$ - $0.2$  mag. for the longer period Cepheids. If this were indeed the case, the same error would apply to PC relations at maximum light. Again, longer period Cepheids obey a flat PC relation at maximum light (KN, SKM, Code 1947), an observation for which there is a sound theoretical argument (see the reference given in Section 5.2). If these longer period Cepheids suffered an extinction error of  $\sim 0.05$ - $0.2$  mag., then the true underlying PC relation at maximum light would be such that as the period increases the colour becomes bluer (that is the temperature is hotter) at maximum light. It is hard to incorporate this in the theoretical framework of Simon et al. (1993), since this paper suggested that for very long period Cepheids ( $\log[P] > 1.5$  roughly), the PC relation at maximum light then starts to have a slope such that Cepheids become redder as the period gets longer. This is in the opposite way to what would need to be true were the PC relation to be linear at mean light. Further, if reddening/extinction errors are indeed making the PC/PL relation appear to be non-linear then the extinction/reddening from two independent methods, as used by the OGLE team and that used in this study, would both need to contain errors of the order of  $\sim 0.05$ - $0.2$  mag. for the longer period Cepheids.

Although longer period Cepheids are generally younger and as such may be surrounded by dust shells which could increase extinction with period and possibly produce a non-linear PC relation as is observed. We argue again that our knowledge of the properties of Cepheids at maximum light as presented above argues against this. Further such an effect would be gradual, however the observed PC relation is sharply non-linear as is evident in the left panel of Figure 3. The use of two extinction maps, one for the OGLE data and another one in this study, to correct for the extinction/reddening that produce similar results also suggest the effect of the extinction is minimal.

The anonymous referee has also suggested the following to test to examine if our extinction corrections do indeed represent individual corrections: plotting the residuals from the  $V$ - and  $R$ -band PL relations against

<sup>7</sup> Some OGLE Cepheids identified by Sebo et al. (2002) are not fundamental mode Cepheids and we exclude them in the table.

**Table 9.** Additional long period Cepheids from Sebo et al. (2002) and Persson et al. (2004).

Name	$\log(P)$	$m_V$	$m_R$	$V$	$R$	$J$	$H$	$K$	$E(B - V)$
Ogle100794	1.0298	17	17	14.78	14.30	...	...	...	0.138
Ogle50505	1.0356	14	13	14.69	14.23	...	...	...	0.127
HV2432	1.0384	7	5	14.24	13.96	12.986	12.653	12.568	0.100
Ogle96614	1.0862	20	19	14.58	14.10	...	...	...	0.138
Ogle117525	1.1016	18	16	14.43	13.92	...	...	...	0.128
Ogle173734	1.1045	6	4	14.72	14.20	...	...	...	0.135
HV2527	1.1122	33	32	14.62	14.16	13.084	12.684	12.570	0.070
HV2260	1.1136	16	14	14.85	14.35	13.263	12.860	12.736	0.130
Ogle270379	1.1233	11	10	14.38	13.85	...	...	...	0.138
HV2579	1.1281	10	7	14.04	13.63	12.688	12.334	12.236	0.100
HV955	1.1377	8	7	14.02	13.62	12.686	12.342	12.251	0.058
HV2538	1.1420	6	3	14.50	14.05	12.864	12.452	12.343	0.100
HV2463	1.1450	10	8	14.22	13.81	12.748	12.381	12.295	0.100
HV5655	1.1526	16	11	14.52	14.06	12.971	12.553	12.443	0.100
Ogle167787	1.1610	11	6	14.09	13.62	...	...	...	0.145
HV12471	1.2001	9	4	14.73	13.95	12.903	12.453	12.308	0.058
Ogle55470	1.2039	9	6	14.65	14.08	...	...	...	0.121
Ogle182466	1.2039	—	—	14.31	14.08	...	...	...	0.127
HV2549	1.2094	9	8	13.67	13.36	12.422	12.078	11.982	0.058
Ogle160625	1.2358	19	14	13.78	13.34	...	...	...	0.138
HV2261	1.2370	11	8	13.26	12.90	...	...	...	0.100
HV2580	1.2285	17	12	14.01	13.53	12.489	12.102	11.999	0.090
HV2836	1.2437	13	9	14.63	14.07	12.747	12.276	12.127	0.180
Ogle250872	1.2709	10	11	13.97	13.52	...	...	...	0.152
U11	1.3026	9	6	13.66	13.23	12.383	11.972	11.860	0.100
Ogle148920	1.3160	21	18	13.60	13.08	...	...	...	0.147
Ogle88441	1.3276	4	4	13.31	13.00	...	...	...	0.142
Ogle286532 <sup>a</sup>	1.3347	12	12	14.01	13.44	...	...	...	0.138
Ogle109640	1.3486	18	16	14.00	13.51	...	...	...	0.138
U1	1.3533	9	5	14.09	13.62	12.350	11.916	11.790	0.100
HV876	1.3561	10	7	13.53	13.32	12.144	11.766	11.665	0.100
HV878	1.3673	33	27	13.49	13.07	12.097	11.735	11.634	0.058
HV938	1.3724	5	5	13.31	12.91	...	...	...	0.100
HV6098	1.3845	8	6	12.95	12.56	11.733	11.405	11.317	0.100
HV902	1.4209	6	4	13.25	12.70	11.833	11.469	11.378	0.070
HV1023	1.4239	22	19	13.77	13.20	12.041	11.618	11.498	0.070
Ogle162232	1.4828	16	11	13.48	12.88	...	...	...	0.152
HV872	1.4750	11	9	13.69	13.15	11.976	11.540	11.422	0.100
HV875	1.4822	16	16	12.96	12.58	11.620	11.285	11.192	0.100
Ogle228645	1.4921	11	7	13.26	12.83	...	...	...	0.129
HV882	1.5027	18	15	13.35	12.90	11.714	11.328	11.202	0.070
HV5761	1.5046	5	3	13.14	12.38	...	...	...	0.100
HV873	1.5359	16	14	13.06	12.58	11.490	11.109	10.998	0.130
HV881	1.5534	16	15	13.09	12.61	11.527	11.142	11.032	0.030
HV2294	1.5626	—	—	...	...	11.237	10.868	10.770	0.070
HV879	1.5656	—	—	13.64	13.01	11.590	11.157	11.031	0.060
HV909	1.5749	—	—	13.03	12.54	11.339	10.984	10.884	0.058
HV2257	1.5943	9	8	13.05	12.53	11.355	10.944	10.833	0.060
HV2338	1.6248	—	—	...	...	11.185	10.798	10.690	0.040
HV877	1.6548	—	—	...	...	11.438	10.962	10.835	0.100
HV900	1.6771	—	—	...	...	11.156	10.751	10.632	0.058
HV953	1.6802	—	—	...	...	10.788	10.423	10.317	0.070
HV2827	1.8969	—	—	...	...	10.428	9.976	9.851	0.080
HV5497	1.9965	9	6	11.93	11.38	10.027	9.602	9.466	0.095
HV2883 <sup>b</sup>	2.0398	—	—	...	...	10.651	10.237	10.098	0.100
HV2447	2.0763	—	—	...	...	10.106	9.704	9.565	0.100
HV883	2.1268	8	6	12.00	11.35	10.302	9.919	9.752	0.100

<sup>a</sup>This Cepheids has relatively low amplitude (Kanbur & Ngeow 2005).<sup>b</sup>This Cepheids has relatively high amplitude (Kanbur & Ngeow 2005).



**Table 10.**  $F$ -test results with additional long period Cepheids, where  $N_{long}^{tot}$  is the total number of long period Cepheids (MACHO + additional Cepheids from Table 9).

Bandpass Criteria	$V$		$R$		$J$		$H$		$K$	
	$N_{long}^{tot}$	$F$	$N_{long}^{tot}$	$F$	$N_{long}^{tot}$	$F$	$N_{long}^{tot}$	$F$	$N_{long}^{tot}$	$F$
1. use all available long period Cepheids	118	7.872	118	8.640	106	12.33	106	10.61	106	2.942
2. exclude Cepheids without $JHK$ means	98	10.96	98	12.45	–	–	–	–	–	–
3. Cepheids with $m_V > 10$	94	8.748	–	–	81	13.71	81	10.29	81	3.849
4. Cepheids with $m_R > 10$	–	–	89	10.32	79	12.70	79	9.603	79	3.681
5. combine item (2), (3) and (4)	79	10.86	79	12.38	79	12.70	79	9.603	79	3.681

each other (see, e.g. Sasselov et al. 1997; Beaulieu et al. 2001). The idea is that if the residuals lie along the reddening line, then the reddening corrections do not represent individual corrections. We carried out such a test with Sample C using the PL relations from a one-line and two-line regression. The results are presented in Figure 7. From the figure it can be seen that the dispersion is reduced when using the one-line regressions as compared to the two-line regression. The points lie along the reddening line.

We, however, argue that this test is not suitable for testing the extinction corrections for the following reasons. First of all, besides the statistical fluctuations, the residuals from the PL relation are expected to originate from three sources (Sasselov et al. 1997): (a) depth dispersion/effects (b) extinction, and (c) intrinsic dispersion due to the width of the instability strip. All three sources will tend to make the residuals be correlated in the  $V$ - and  $R$ -bands, as seen in Figure 7. We ignore the discussion of the depth effect (represented as dashed lines in Figure 7) because we assume the depth effect is negligible for the LMC Cepheids. For the source from extinction, recall that the extinction correction for a Cepheid is calculated using  $A_\lambda = R_\lambda E(B - V)$ , where the  $E(B - V)$  value is either the mean value of the LMC or an estimated value for individual Cepheids. Hence a correlation of the residuals is naturally introduced (represented as solid lines in Figure 7) for the reddening vector used in this paper (Section 2), because  $R_\lambda/R_V \simeq \text{constant}$ . Even for a hypothetical group of Cepheids that are free from extinction and depth dispersion, the residuals of the  $V$ - and  $R$ -band PL relations will still be correlated. This is due to the existence of the instability strip and the PLC relation (Sandage 1958; Madore & Freedman 1991). Therefore mapping the Cepheids from the instability strip (along the constant-period line) to the  $V$ - and  $R$ -band PL relations will cause the residuals to be unavoidably correlated. Unfortunately, this line of residuals is very close to the reddening line in the residual plots (not shown in Figure 7, see also Sasselov et al. 1997). This makes it hard to say whether the correlation of the residuals seen in Figure 7 is caused by the extinction or the width of the instability strip<sup>8</sup>.

<sup>8</sup> The spread of the residuals ( $\sim \pm 1.0$  mag.) is comparable to the findings of Sasselov et al. (1997) and Beaulieu et al. (2001). This is not a surprise because: (a) Cepheids in the observed instability strip are roughly normally distributed (with  $FWHM \sim \sigma_{IS}$  and dominate the overall distribution); and

In this paper we have tried our best to correct for the extinction of the LMC Cepheids. Based on the above arguments we believe that extinction errors are not the physical cause for the observed non-linear LMC PL relations. Other physical reasons, such as the pulsational properties (Bono et al. 1999; Caputo et al. 2000) and/or the internal structure (Simon et al. 1993; Kanbur et al. 2004; Kanbur & Ngeow 2005) of the Cepheid variables (we refer to these as “internal reasons” as opposed to the “external reasons” like dust extinction) may be responsible for the observed non-linear PL relations. We have provided compelling evidence that the (extinction corrected) Cepheid PL relation in the LMC is non-linear around a period of 10 days in the optical  $V$ - &  $R$ -bands. Assuming the random phase corrections methods used in this study, we also find the PL relation in the  $J$ - &  $H$ -band to be non-linear around this period but that the  $K$ -band relation is marginally linear. An investigation into the physics behind this is left for future studies.

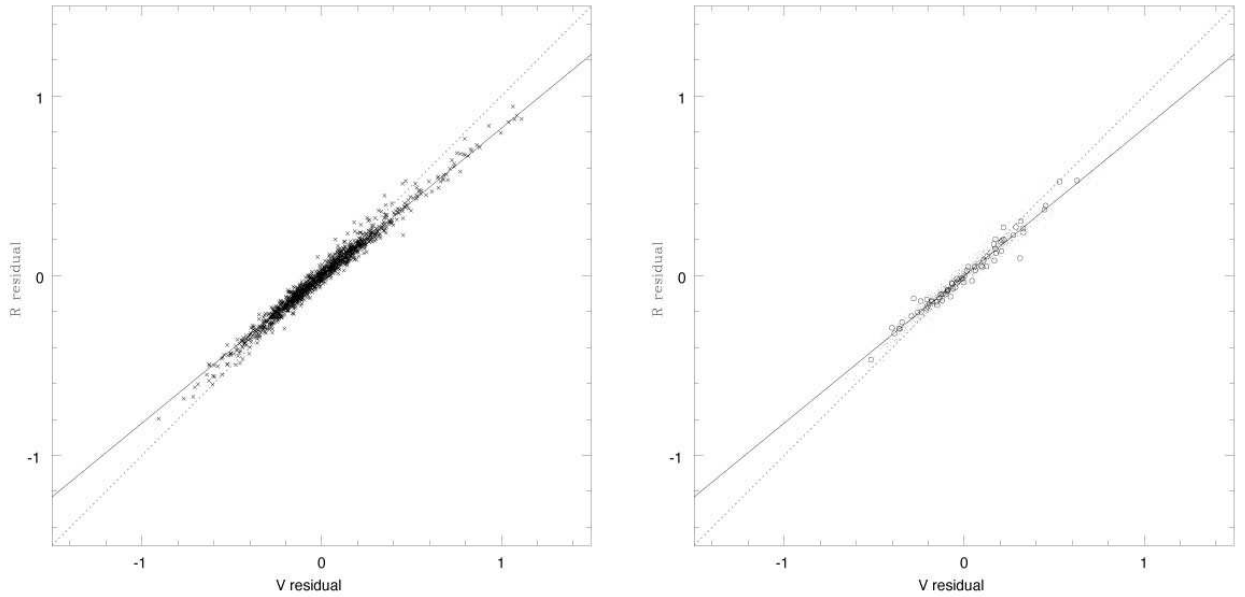
## ACKNOWLEDGEMENTS

The authors would like to thank the assistance from D. Zaritsky and J. Harris regarding the usage of LMC extinction map. SN’s & KC’s works were performed under the auspices of the U.S. Department of Energy, National Nuclear Security Administration by the University of California, Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48. This publication makes use of data products from the Two Micron All Sky survey (2MASS), which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center, funded by the National Aeronautics and Space Administration and the National Science Foundation. The authors would also like to thank an anonymous referee for several comments which helped to improve the manuscript.

## REFERENCES

Andreasen, G., 1988, *A&A*, 196, 159

(b) the measurement and extinction errors are also roughly normally distributed. Hence the overall distribution is also normally distributed with tails extended to  $\sim \pm 1.0$  mag. at both ends.



**Figure 7.** The residual plots of the  $V$ - and  $R$ -band PL relations from one-line regression (left panel) and two-line regression with a break at 10 days (right panel). In the right panel, the dots and open circles represent the residuals from the short (less than 10 days) and long period Cepheids, respectively. The solid and dashed lines represent the reddening lines (defined by our reddening vector) and the expected residuals due to depth dispersion (diagonal lines), respectively.

Alcock, C., Allsman, R., Alves, D., Axelrod, T., Becker, A., Bennett, D., Bersier, D., Cook, K., et al., 1999, *AJ*, 117, 920  
 Alcock, C., Allsman, R., Alves, D., Axelrod, T., Becker, A., Bennett, D., Cook, K., Dalal, N., et al., 2000, *ApJ*, 542, 281  
 Baraffe, I., Alibert, Y., Méra, D., Chabrier, G. & Beaulieu, J.-P., 1998, *ApJ*, 499, L205  
 Baraffe, I. & Alibert, Y., 2001, *A&A*, 371, 592  
 Bauer, F., Afonso, C., Albert, J. N., Alard, C., Andersen, J., Ansari, R., Aubourg, E., Bareyre, P., et al., 1999, *A&A*, 348, 175  
 Beaulieu, J.-P., Buchler, R. & Kolláth, Z., 2001, *A&A*, 373, 164  
 Bersier, D., Burki, G. & Burnet, M., 1994, *A&AS*, 108, 9  
 Bono, G., Caputo, F. & Marconi, M., 1998, *ApJ*, 497, L43  
 Bono, G., Caputo, F., Castellani, V. & Marconi, M., 1999, *ApJ*, 512, 711  
 Bono, G., Caputo, F., Cassisi, S., Marconi, M., Pier-santi, L. & Tornambé, A., 2000a, *ApJ*, 543, 955  
 Bono, G., Marconi, M. & Stellingwerf, R., 2000b, *A&A*, 360, 245  
 Caputo, F., Marconi, M. & Ripepi, V., 1999, *ApJ*, 525, 784  
 Caputo, F., Marconi, M. & Musella, I., 2000, *A&A*, 354, 610  
 Coulson, I., Caldwell, J. & Gieren, W., 1985, *ApJS*, 57, 595  
 Cleveland, W. S., 1979, *Journal of the American Statistical Association*, 74, 829  
 Code, A. D., 1947, *ApJ*, 106, 309

Cox, J., 1980, *Theory of Stellar Pulsation*, Princeton University Press, 1<sup>st</sup> Ed.  
 Feigelson, E. & Babu, G. J., 1992, *ApJ*, 397, 55  
 Fernie, J. D., 1967, *AJ*, 72, 1327  
 Fernie, J. D., 1969, *PASP*, 81, 707  
 Fisher, M. A., Bolles, R. C., 1981, *Comm. of the ACM*, 24, 381  
 Freedman, W., Madore, B., Gibson, B., Ferrarese, L., Kelson, D., Sakai, S., Mould, J., Kennicutt, R., et al., 2001, *ApJ*, 553, 47  
 Gieren, W., Fouqué, P. & Gómez, M., 1998, *ApJ*, 496, 17  
 Gieren, W., Moffett, T. & Barnes, T., 1999, *ApJ*, 512, 553  
 Gieren, W., Storm, J., Barnes, T., Fouqué, P., Pietrzyński, G. & Kienzie, F., 2005, *ApJ*, 627, 224  
 Isobe, T., Feigelson, E., Akritas, M. & Babu, G. J., 1990, *ApJ*, 364, 104  
 Kanbur, S., Iono, D., Tanvir, N. & Hendry, M., 2002, *MNRAS*, 329, 126  
 Kanbur, S., Ngeow, C., Nikolaev, S., Tanvir, N. & Hendry, M., 2003, *A&A*, 411, 361  
 Kanbur, S. & Ngeow, C., 2004, *MNRAS*, 350, 962 (KN)  
 Kanbur, S., Ngeow, C. & Buchler, R., 2004, *MNRAS*, 354, 212  
 Kanbur, S. & Ngeow, C., 2005, *MNRAS* submitted  
 Laney, C. & Stobie, R., 1994, *MNRAS*, 266, 441  
 Leonard, D., Kanbur, S., Ngeow, C. & Tanvir, N., 2003, *ApJ*, 594, 247  
 Madore, B., 1985, *ApJ*, 298, 340  
 Madore, B. & Freedman, W., 1991, *PASP*, 103, 933  
 Marconi, M., Musella, I. & Fiorentino, G., 2005, *ApJ* accepted (astro-ph/0506207)

- Moffett, T., Gieren, W., Barnes, T. & Gómez, M., 1998, ApJS, 117, 135
- Ngeow, C., Kanbur, S., Nikolaev, S., Tanvir, N. & Hendry, M., 2003, ApJ, 586, 959
- Ngeow, C. & Kanbur, S., 2005, MNRAS, 360, 1033
- Nikolaev, S., Drake, A., Keller, S., Cook, K., Dalal, N., Griest, K., Welch, D. & Kanbur, S., 2004, ApJ, 601, 260
- Persson, S., Madore, B., Krzemiński, W., Freedman, W., Roth, M. & Murphy, D., 2004, AJ, 128, 2239
- Press, W., Teukolsky, S., Vetterling, W. & Flannery, B., 1992, *Numerical Recipes in C*, Cambridge University Press, 2<sup>nd</sup> ed.
- Riess, A., Li, W., Stetson, P., Filippenko, A., Jha, S., Kirshner, R., Challis, P., Garnavich, P. & Chornock, R., 2005, ApJ, 627, 579
- Saha, A., Sandage, A., Tammann, G., Dolphin, A., Christensen, J., Panagia, N. & Macchetto, F., 2001, ApJ, 562, 314
- Sandage, A., 1958, ApJ, 127, 513
- Sandage, A. & Tammann, G. A., 1968, ApJ, 151, 531
- Sandage, A., Tammann, G. A. & Reindl, B., 2004, A&A, 424, 43
- Sasselov, D., Beaulieu, J.-P., Renault, C., Grison, P., Ferlet, R., Vidal-Madjar, A., Maurice, E., Prévot, L., et al., 1997, A&A, 324, 471
- Sebo, K., Rawson, D., Mould, J., Madore, B., Putman, M., Graham, J., Freedman, W., Gibson, B. & Germany, L., 2002, ApJS, 142, 71
- Shobbrook, R., 1992, MNRAS, 255, 486
- Simon, N. & Lee, A., 1981, ApJ, 248, 291
- Simon, N. & Moffett, T., 1985, PASP, 97, 1078
- Simon, N., Kanbur, S. & Mihalas, D., 1993, ApJ, 414, 310
- Skrutskie, M., 1998, in *The Impact of Near-Infrared Sky Surveys on Galactic and Extra-galactic Astronomy*, eds. N. Epchtein, Astrophysics & Space Science Library Vol. 230, Kluwer Academic Pub., Dordrecht, pg. 11
- Storkey, A. J., Hambly, N. C., Williams, C. K. I. & Mann, R. G., 2004, MNRAS, 347, 36
- Tammann, G. A. & Reindl, B., 2002, Ap&SS, 280, 165
- Tammann, G. A., Reindl, B., Thim, F., Saha, A. & Sandage, A., 2002, in *A New Era in Cosmology*, eds. N. Metcalfe & T. Shanks, ASP Conf. Series Vol. 283, Astron. Soc. Pac., San Francisco, pg. 258
- Tanvir, N., 1997, in *The Extra-galactic Distance Scale*, STScI May Symposium, eds. Livio, Donahue & Panagia, Cambridge University Press, pg. 91
- Thim, F., Tammann, G., Saha, A., Dolphin, A., Sandage, A., Tolstoy, E. & Labhardt, L., 2003, ApJ, 590, 256
- Thim, F., Hoessel, J., Saha, A., Claver, J., Dolphin, A. & Tammann, G., 2004, AJ, 127, 2322
- Udalski, A., Szymański, M., Kubiak, M., Pietrzyński, G., Soszyński, I., Wozniak, P. & Zebrun, K., 1999a, AcA, 49, 201
- Udalski, A., Soszyński, I., Szymański, M., Kubiak, M., Pietrzyński, G., Wozniak, P. & Zebrun, K., 1999b, AcA, 49, 223
- Weisberg, S., 1980, *Applied Linear Regression*, John Wiley & Sons, 1<sup>st</sup> Ed.
- Zaritsky, D., Harris, J., Thompson, I. & Grebel, E., 2004, AJ, 128, 1606